

ASYMPTOTIC INFERENCE FOR LOCALLY STATIONARY PROCESSES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Inder Rafael Tecuapetla Gómez

August 2013

© 2013 Inder Rafael Tecuapetla Gómez

ALL RIGHTS RESERVED

ASYMPTOTIC INFERENCE FOR LOCALLY STATIONARY PROCESSES

Inder Rafael Tecuapetla Gómez, Ph.D.

Cornell University 2013

The study of locally stationary processes contains theory and methods about a class of processes that describe random phenomena whose fluctuations occur both in time and space.

We consider three aspects of locally stationary processes that have not been explored in the already vast literature on these nonstationary processes.

We begin by studying the asymptotic efficiency of simple hypotheses tests via large deviation principles. We establish the analogues of classic results such as Stein's lemma, Chernoff bound and the more general Hoeffding bound. These results are based on a large deviation principle for the log-likelihood ratio test statistic between two locally stationary Gaussian processes which is obtained and presented in the first chapter.

In the second chapter we consider the Bayesian estimation of two parameters of a locally stationary process: trend and time-varying spectral density functions, respectively. Under smoothness conditions on the latter function, we obtain the asymptotic normality and efficiency, with respect to a broad class of loss functions, of Bayesian estimators. In passing we also show the asymptotic equivalence between Bayesian estimators and the maximum likelihood estimate.

Our concluding fourth chapter explores the time-varying spectral density estimation problem from the point of view of Le Cam's theory of statistical experiments. We establish that the estimation of a time-varying spectral density function can be asymptotically construed as a white noise problem with drift. This result is based on Le Cam's connection theorem.

BIOGRAPHICAL SKETCH

Inder was born and raised in Campeche, México the 16th of September of 1981. In 1999 he joined the Department of Mathematics at the Universidad de Guanajuato as an undergraduate. In 2004 he defended the thesis *Characterization of the Logit model* and obtained a B.Sc. from that institution.

Also in Guanajuato, Inder pursued Master's studies at the Center for Mathematical Research (CIMAT) and in 2006 he defended the thesis *Extremal index in heteroscedastic processes* as part of the requirements to gain an M.Sc. specializing in Probability and Statistics.

In the summer of 2007, Inder left México and joined Cornell University. Since 2010 he has been collaborating with Professor Nussbaum on several projects. In May 2013, some of those projects were presented as part of his Ph.D. dissertation.

To Rosalba and Gibrán (of course)

ACKNOWLEDGEMENTS

There are several people who are, to different extents, responsible for this manuscript. I only hope that my vague memory does not betray me now at the very moment of saying *thank you*.

I commence this section by recognizing the patience that Professor Nussbaum has had with my erratic wanders into the world of mathematical statistics. Without his outstanding mentoring this manuscript would certainly not have been possible.

During my most fateful time in the Ph.D. program Professor Wells knew how to make me feel part of the department and even a small conversation with him was sufficient to encourage me to continue with my investigations. Professor Resnick is partially responsible for my incursion in the world of *applied statistics* and I will always be grateful to him.

The Ph.D. program has given me many opportunities. One of the most significant is perhaps the great chance that I had to coincide, in lovely Ithaca, with an incredible group of people, most of them are also Cornell Ph.D. students (Ryan McQuinn, Rafael Orozco, Kalyani Raghunathan, Joyjit Roy), some of them are students in the Department of Statistics (Irina Gaynanova, Will Nicholson, Jón Árne Steingrímsson, Leifur Thorbergsson), only a few were part of the Linux lab crew (Michael Grabchak, Matthias Kormaksson, Rajendran Narayanan and Dave Zeber), but certainly all of *you* are my very esteemed friends.

I would like to express all my gratitude to my mother whose continuous sacrifice has always been my inspiration to try and get the most out of every project that I join.

The beginning of my doctoral studies were sponsored by the Mexican Council of Science and Technology (CONACyT) through their grant 230788. The last 3 years I was financially supported by the National Science Foundation through the grant DMS-1106460. I am grateful to these scientific organizations.

TABLE OF CONTENTS

1	Introduction	1
1.1	Motivation	1
1.2	Basic definitions	3
1.3	Examples	6
1.4	Organization of dissertation	9
2	On large deviations for testing simple hypotheses for locally stationary processes	11
2.1	Introduction and main results	11
2.2	Large deviations for locally stationary Gaussian processes	14
2.2.1	A comment on large deviations for non-Gaussian locally stationary processes	21
2.3	Proof of Stein's lemma	24
2.4	Proof of Chernoff bound	27
2.5	Appendix	28
3	Bayesian estimators for locally stationary Gaussian processes	31
3.1	Introduction	31
3.2	Main results	34
3.2.1	Bayesian estimators	38
3.2.2	Maximum likelihood estimator	43
3.3	Appendix	44
4	Global parametric asymptotic equivalence of locally stationary processes	57
4.1	Introduction and main result	57
4.2	From local to global equivalence: LeCam's connection theorem	59
4.3	Local asymptotic equivalence	64
4.3.1	Local asymptotic equivalence of lsGp	65
4.3.2	Local asymptotic equivalence of a bidimensional white noise model	67
4.4	Preliminary estimators	72
4.4.1	Estimators in a bidimensional white noise model	73
4.5	Strong local convergence of statistical experiments	76
	Bibliography	79

LIST OF FIGURES

1.1	Here $\mu_t = \exp\{t^4 + \cos(t)\}$ and $\sigma_t = \log(t^2)$	6
1.2	Time-varying spectral density of a tv-AR(2) with $\sigma(u) = \sin(2\pi u)$, $a_0(u) = 1$, $a_1(u) = u/2$, $a_2(u) = 0.1 + 0.3u^2$	8
1.3	From top to bottom: time-varying spectral density of a tv-AR(2) at time-points $u = \{0.2, 0.4, 0.75\}$	8

CHAPTER 1

INTRODUCTION

1.1 Motivation

The study of stationary processes has historically received a great deal of attention from researchers of several scientific fields. There exists a vast literature on theory and methods for this type of processes. Sometimes this methodology is applied even in situations where it is clear that a non stationary model is more appropriate; an example of this is when a time series shows evidence of instability in its stochastic characteristics, such as time-varying trends or evolutionary second order structure.

To overcome the restrictions imposed by the assumption of stationarity, Rainer Dahlhaus [19] developed a theory and method to study locally stationary processes, a class of non stationary processes. The main feature of this class of processes is that its spectral behavior is allowed to change slowly over time. Dahlhaus' approach allows for an asymptotic treatment of statistical inference problems of a class of non stationary processes, a feature which had not been provided by previous efforts to set a meaningful framework to study non stationarity.

The seminal work of Dahlhaus has been applied to diverse scientific fields and extended in different manners. At this point, we only mention that recently models based on locally stationary processes have been implemented in fields such as seismology, medicine and computer science, cf. Beran [4], Eckley et al. [25] and Maharaj and Alonso [37]. Additionally, processes with long-range dependence have been studied within the framework of local stationarity, cf. Palma [41] and Ferreira et al. [26] among others.

In this work we are interested in performing asymptotic statistical inference on locally stationary processes. To be more precise, we combine theories from applied probability and asymptotic estimation to study locally stationary Gaussian processes from three different angles associated with general asymptotic inference: *efficiency of tests*, *Bayesian estimators* and *decision theory*.

We may say that our work is another story about the likelihood ratio statistic, but a particular one, that of the likelihood ratio statistic of locally stationary Gaussian processes. Indeed, most of our results rely on suitable expressions of the likelihood ratio process associated with the probability law of a locally stationary Gaussian process. Under certain smoothness conditions, this likelihood ratio will be shown to be sufficiently regular and we will apply known results from theories such as *large deviations*, *asymptotic Bayesian estimation* and *LeCam theory of statistical experiments* to obtain our main results.

Even though Gaussianity greatly simplifies our calculations, most of our proofs will be long and tedious. Some of our results, however, may be easily extended to non Gaussian cases or long-dependent process. Patience and serenity are required to succeed in this future task.

The remaining of this introductory chapter is structured as follows. In Section 1.2 we present a proper definition of the class of locally stationary Gaussian processes as well as the main features of the time-varying spectral density function. In particular, the asymptotic relation between this function and the covariance matrix of a locally stationary process is emphasized. Section 1.3 contains some examples of useful models which belong to the class of locally stationary processes; time-varying-ARMA(p, q) processes are locally stationary. In the concluding Section 1.4 we briefly explain the content of each of the following chapters of this work.

1.2 Basic definitions

We begin by defining locally stationary Gaussian processes. Let $\mathbb{T} = [0, 1] \times [-\pi, \pi]$.

Definition 1.1. A triangular array of random variables $X := (X_{k,n})_{1 \leq k \leq n}$ is said to be a *locally stationary Gaussian process* with trend function μ and transfer function A° if the following spectral representation is valid:

$$X_{k,n} = \mu\left(\frac{k}{n}\right) + \int_{-\pi}^{\pi} \exp\{i k \lambda\} A_{k,n}^\circ(\lambda) d\mathbf{B}(\lambda), \quad (1.1)$$

where the stochastic process \mathbf{B} is a Brownian motion on $[-\pi, \pi]$. Additionally, we suppose that there exist a constant C and an *approximating function* $A : \mathbb{T} \rightarrow \mathbb{C}$ satisfying

$$\sup_{k,\lambda} |A_{k,n}^\circ(\lambda) - A(k/n, \lambda)| \leq \frac{C}{n}. \quad (1.2)$$

For a definition of locally stationary processes which are not necessarily Gaussian we refer to Dahlhaus [19] and more recently to Hirukawa and Taniguchi [30].

Notice that when μ and A° do not depend on k and n then X is independent on n as well and the standard spectral representation of a stationary sequence is obtained. That is, the theory for locally stationary processes generalizes the classical theory for stationary processes. Although in practice we only observe one realization of a locally stationary process, by re-scaling the number of observations, from $\{1, \dots, n\}$ to $\{1/n, 2/n, \dots, 1\}$, when n is large enough, we have a framework in which there is more information about the process on each interval of length $1/n$. Thus this approach allows us to develop meaningful statistical inference about X in despite of its intrinsic nonstationary nature.

Representations similar to (1.1) were considered by Priestley in his theory of evolutionary spectra; cf. [42]. Notice that by including the approximating function A in (1.2),

a considerable amount of nonstationary models can be studied under the framework of locally stationary processes. For instance, time-varying ARMA(p, q) models satisfy (1.1). We refer to Dahlhaus [20] for a more detailed comparison between the theories of evolutionary spectra and locally stationary processes.

Associated with the approximating function A we have the time-varying spectral density function $f : \mathbb{T} \rightarrow \mathbb{R}$ which is defined as

$$f(u, \lambda) := (2\pi)^{-1} |A(u, \lambda)|^2. \quad (1.3)$$

Although f , as defined above, may not be unique, if the approximating function A is sufficiently smooth, then the function f is asymptotically uniquely determined by the whole triangular array $(X_{k,n})_{1 \leq k \leq n}$. Indeed, if A is uniformly Lipschitz continuous in both arguments with index $\alpha > 1/2$ then for all $u \in (0, 1)$,

$$\int_{-\pi}^{\pi} |f_n(u, \lambda) - f(u, \lambda)|^2 d\lambda = o(1),$$

where $f_n(u, \lambda)$ is the Wigner-Ville spectrum of X for fixed n :

$$f_n(u, \lambda) := \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \text{cov}(X_{[un+s/2],n}; X_{[un-s/2],n}) \exp(-i\lambda s),$$

where $X_{s,n}$ obeys the representation (1.1) and $[x]$ denotes then largest integer less or equal than x . That is, for each time-point $u \in (0, 1)$, f is the $\mathbb{L}^2(-\pi, \pi)$ -limit of the Wigner-Ville spectrum f_n . Thus, the function f can be construed as the spectral density of a process which smoothly deviates from a stationary framework. This result is Theorem 2.2 of Dahlhaus [18].

Another useful element for this work is the time-varying covariance function at time-point $u \in [0, 1]$ and lag $k \in \mathbb{Z}$. More precisely,

$$c(u, k) = \int_{-\pi}^{\pi} f(u, \lambda) \exp\{-i k \lambda\} d\lambda. \quad (1.4)$$

For two transfer functions A° and B° , set the $n \times n$ matrix

$$\Sigma_n(A, B) := \left(\int_{-\pi}^{\pi} \exp\{i(k-l)\lambda\} A_{k,n}^\circ(\lambda) \overline{B_{l,n}^\circ(\lambda)} d\lambda \right)_{k,l=1,\dots,n}. \quad (1.5)$$

If the true process is locally stationary with transfer function A° then $\Sigma_n = \Sigma_n(A, A)$ is the true (and theoretical) covariance matrix of the process X .

Assuming that the approximating function A is uniformly continuous in u it is known that the (r, t) -th element of the theoretical covariance matrix Σ_n of a locally stationary process satisfies

$$\text{cov}(X_{r,n}; X_{r+(t-r),n}) = c\left(\frac{r}{n}, t-r\right) + \varepsilon_n, \quad (1.6)$$

where $\varepsilon_n = O(n^{-1})$; cf. Remark 2.8 of Dahlhaus [21].

In view of these results, instead of dealing with the theoretical covariance matrix of a locally stationary process, we can appropriately approximate it if we use a matrix whose entries are calculated using the time-varying covariance function. Observe that the latter function can be construed as the Fourier coefficient of $f(u, \cdot)$ for each time-point $u \in (0, 1)$. Thus, the asymptotic relation between the time-varying spectral density function and the theoretical covariance of a locally stationary process X is a natural generalization of that between spectral density function and covariance matrix of a stationary process.

In summary, if X is the realization of a locally stationary Gaussian process with trend function μ and time-varying spectral density function f , then X can be construed as a realization of an n -dimensional Gaussian random variable with mean μ and covariance matrix Σ_n . The latter matrix is given by eq. (1.5) and from eq. (1.6) its can be approximated through the values of f .

1.3 Examples

At this point it is convenient to show examples of objects which satisfy the definition and conditions that we have introduced in previous sections. Here we have two examples of processes which are locally stationary.

1. If Y_t is a stationary process with spectral representation

$$Y_t = \int_{-\pi}^{\pi} \exp\{-i \lambda t\} A(\lambda) d\varepsilon(\lambda)$$

and $\mu, \sigma : [0, 1] \rightarrow \mathbb{R}$ are continuous functions, then

$$X_t = \mu(t/n) + \sigma(t/n)Y_t$$

is locally stationary. Indeed, observe that Definition 1.1 and eq. 1.2 are satisfied with $A_{k,n}^{\circ}(\lambda) = A(k/n, \lambda) = \sigma(k/n) A(\lambda)$.

A possible realization of the process X_t is presented below:

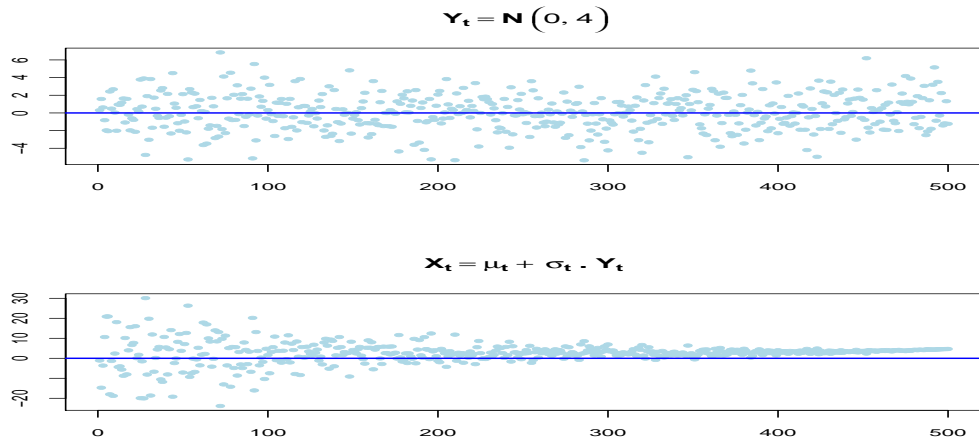


Figure 1.1: Here $\mu_t = \exp\{t^4 + \cos(t)\}$ and $\sigma_t = \log(t^2)$.

2. Consider the system of difference equations:

$$\sum_{j=0}^p a_j \left(\frac{k}{n} \right) \left(X_{k-j,n} - \mu \left(\frac{k-j}{n} \right) \right) = \sum_{i=0}^q b_i \left(\frac{k}{n} \right) \sigma \left(\frac{k-i}{n} \right) \varepsilon_{k-i}, \quad k \in \mathbb{Z}, \quad (1.7)$$

where the ε_k are independent random variables with $E\varepsilon_k = 0$, $E|\varepsilon_k| < \infty$, $a_0(u) \equiv b_0(u) \equiv 1$ and $a_j(u) = a_j(0) = b_i(u) = b_i(0)$, $u < 0$. Observe that if the functions $a_j(u) = a$ and $b_i(u) = b$ for all u , then the system of equations given above corresponds to an ARMA(p, q) process.

If the functions $a_j(\cdot)$, $b_i(\cdot)$ and $\sigma^2(\cdot)$ are of bounded variation and

$$\sum_{j=0}^p a_j(u) z^j \neq 0$$

for all u and all $0 < |z| \leq 1 + \delta$ for some $\delta > 0$ then the system of equations (1.7) has a solution which satisfies Definition 1.1 and eq. 1.2 with

$$A_{k,n}^\circ(\lambda) = \frac{1}{\sqrt{\pi}} \sum_{l=0}^{\infty} \phi_{k,n,l} \sigma(k/n) \exp\{-i \lambda l\},$$

$$A(k/n, \lambda) = \frac{\sigma(k/n)}{\sqrt{2\pi}} \frac{1 + \sum_{i=1}^q b_i(u) \exp(i \lambda i)}{1 + \sum_{j=1}^p a_j(u) \exp(i \lambda j)}.$$

for some $\phi_{k,n,l}$ where $\sum_{l=0}^{\infty} |\phi_{k,n,l}| < \infty$. We refer to Proposition 2.3 of Dahlhaus and Polonik [22] for a proof of this result.

This result shows that time-varying ARMA(p, q) processes are locally stationary with time-varying spectral density function given by

$$f(u, \lambda) = \frac{\sigma_\theta^2(u)}{2\pi} \frac{\left| \sum_{i=0}^q b_i(u) \exp(i \lambda i) \right|^2}{\left| \sum_{j=0}^p a_j(u) \exp(i \lambda j) \right|^2}.$$

Figure 2 shows a graphical representation of the time-varying spectral density function of a tv-AR(2) processes. Specifically, we consider a time-varying AR(2) process where we have chosen the following parameters: $\sigma(u) = \sin(2\pi u)$, $a_0(u) = 1$, $a_1(u) = u/2$ and $a_2(u) = 0.1 + 0.3u^2$.

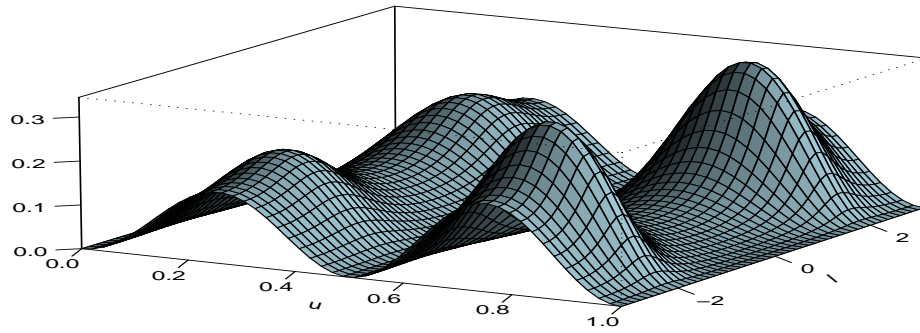


Figure 1.2: Time-varying spectral density of a tv-AR(2) with $\sigma(u) = \sin(2\pi u)$, $a_0(u) = 1$, $a_1(u) = u/2$, $a_2(u) = 0.1 + 0.3u^2$.

This figure shows the oscillating behavior of a time-varying spectral density; depending upon the time-point $u \in (0, 1)$ that we consider, the appearance of function $f(u, \cdot)$ will change.

This smooth change can be better distinguished in Figure 2 where we have selected three time-points, to wit, $u = 0.2$, $u = 0.4$ and $u = 0.75$, and plot the corresponding spectral densities $f(u, \lambda)$ for each value of u .

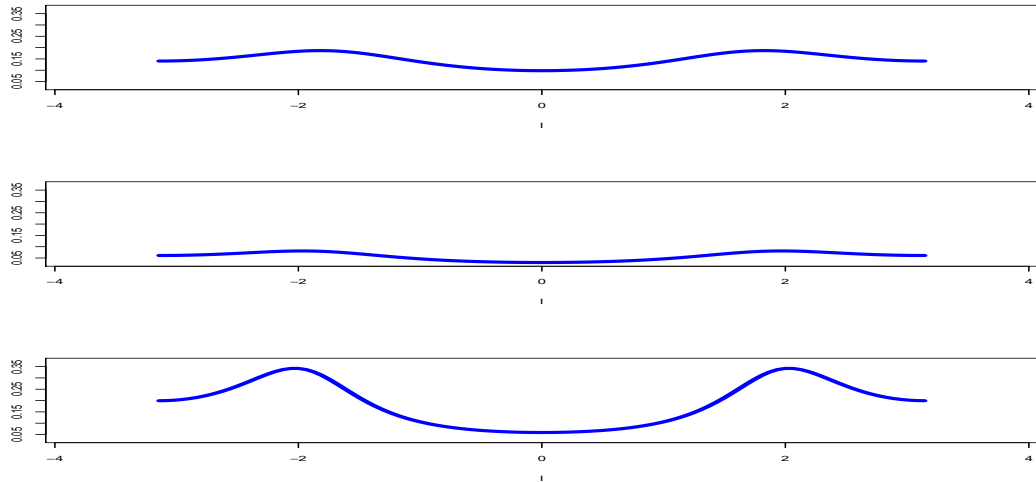


Figure 1.3: From top to bottom: time-varying spectral density of a tv-AR(2) at time-points $u = \{0.2, 0.4, 0.75\}$.

1.4 Organization of dissertation

Having presented the basic notions of locally stationary processes in this introductory chapter, now we are ready to elaborate in the contributions of our work. In the following chapters we will benefit from specific areas of applied probability and asymptotic estimation to provide some results on the asymptotic behavior of statistical procedures related with locally stationary processes.

Chapter 2 presents *large deviations for testing simple hypotheses for locally stationary processes*. In this chapter we derive a large deviation principle for the log-likelihood ratio statistic between two locally stationary processes. We apply this result to obtain the exponential rates of convergence of Type-I and Type-II error probabilities in a simple hypothesis testing problem. Further application of our large deviation result allows us to establish the asymptotic Kullback-Leibler divergence between two locally stationary processes.

Chapter 3 is about *Bayesian estimators for locally stationary Gaussian processes*. This chapter is devoted to analyze the likelihood process when its parameters are properly localized. In particular, we show the uniform local asymptotic normality of this process over compact sets of \mathbb{R}^d . This result will allow us to prove consistency, asymptotic normality as well as efficiency of a broad class of Bayesian estimators of the trend and time-varying spectral density functions.

In *global parametric asymptotic equivalence for locally stationary Gaussian processes*, our concluding Chapter 4, we focus on the estimation of the time-varying spectral density function from the viewpoint of LeCam's theory of statistical experiments.

We utilize the uniform local asymptotic normality of the likelihood ratio process to show that asymptotically the problem of estimating the time-varying spectral density function can be construed as a white noise problem with drift. Our result is based on the LeCam's connection theorem, a general result which allows us to compare statistical experiments when these have parameter spaces with dimensionality restrictions.

In each of these three chapters we comment on possible extensions of our results to frameworks where Gaussianity is no longer assumed or where the estimation problem is purely nonparametric.

CHAPTER 2

ON LARGE DEVIATIONS FOR TESTING SIMPLE HYPOTHESES FOR LOCALLY STATIONARY PROCESSES

In this chapter we derive a large deviation result for the log-likelihood ratio for testing simple hypotheses in locally stationary Gaussian processes. This result allows us to find explicitly the rates of exponential decay of the error probabilities of type I and type II for Neyman-Pearson tests. Furthermore, we obtain the analogue of classical results on asymptotic efficiency of tests such as Stein's lemma and the Chernoff bound, as well as the more general Hoeffding bound concerning best possible joint exponential rates for the two error probabilities.

2.1 Introduction and main results

Consider a locally stationary Gaussian process $X := (X_{k,n})_{1 \leq k \leq n}$ and the problem of testing

$$H_0 : X \sim \mathbf{P}_n \quad \text{vs} \quad H_1 : X \sim \mathbf{Q}_n, \quad (2.1)$$

where $\mathbf{P}_n \sim \mathcal{N}_n(0, \Sigma_n)$ and $\mathbf{Q}_n \sim \mathcal{N}_n(0, \widetilde{\Sigma}_n)$. In our setting, both hypotheses remain fixed as $n \rightarrow \infty$, so the study of error probabilities has to be based on large deviation theory. The main objective of this chapter is to characterize the exponential rates of decrease for the error probabilities of the first and second kind and their interdependence. In the information theoretical literature, this characterization (for the i.i.d. case) is known as the Hoeffding bound (Hoeffding [31], Csiszár and Longo [14] and Blahut [7]); Stein's lemma and the Chernoff bound then appear as special cases. For a concise presentation of these results in the context of large deviation theory and asymptotic statistics see Genon-Catalot and Picard [28]. In order to treat the case of locally stationary Gaus-

sian processes we mainly follow some work by Bouaziz [8] and Taniguchi and Kakizawa [46] who have obtained large deviations for stationary Gaussian sequences.

Throughout this chapter $X := (X_{k,n})_{1 \leq k \leq n}$ will denote observations from a locally stationary Gaussian process with zero trend function, transfer function A° , approximating function A and time-varying spectral density f . We study asymptotic statistical inference about X under the two simple hypotheses established in (2.1).

In accordance with the results presented in the introductory chapter under H_0 , the covariance matrix, Σ_n , is asymptotically defined through the time-varying spectral density f . Under H_1 , the covariance matrix of X , denoted by $\widetilde{\Sigma}_n$, is asymptotically defined via another time-varying spectral density, say g .

Let T_n be a test for the hypotheses in (2.1). The performance of T_n is determined by the false-alarm rate (type I error probability or simply α_n) and the nondetection rate (type II error probability or just β_n) which by definition are

$$\alpha_n = \mathbf{P}\{T_n \text{ Rejects } H_0 \mid H_0\}, \quad \beta_n = \mathbf{P}\{T_n \text{ Rejects } H_1 \mid H_1\}. \quad (2.2)$$

We are interested in obtaining exponential rates of decrease of α_n and β_n for Neyman-Pearson and Bayes-type tests. In order to formulate our main results some notation is in order. Set $\mathbb{T} = [0, 1] \times [-\pi, \pi]$. Let $\mathcal{B}_{\mathbb{T}}$ denote the set of functions $h : \mathbb{T} \rightarrow \mathbb{C}$ such that:

1. There exists a constant $C > 0$ such that $|h(u, \lambda)| > C$, for all $(u, \lambda) \in \mathbb{T}$, and
2. the derivative $\frac{\partial}{\partial u} \frac{\partial}{\partial \lambda} h$ is uniformly bounded.

Proposition 2.1. *[Stein's lemma for LSP] Suppose that $f, g \in \mathcal{B}_{\mathbb{T}}$ and $f \neq g$ on \mathbb{T} . Let β_n^ε be the infimum of β_n among all tests with $\alpha_n < \varepsilon$. Then for any $\varepsilon < 1$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^\varepsilon = -K_L(f; g),$$

where

$$K_L(f; g) = \frac{1}{4\pi} \int_{\mathbb{T}} \left[\log \frac{g}{f} + \frac{f - g}{g} \right] d\lambda du.$$

Observe that $f := f(u, \lambda)$ and $g := g(u, \lambda)$.

In the scenario of having some *a priori* probability on H_0 we have the following result.

Proposition 2.2. *[Chernoff bound for LSP] Suppose that the conditions of Proposition 1 hold. Let $\mathbb{P}_n^B = \alpha_n \mathbb{P}\{H_0\} + \beta_n \mathbb{P}\{H_1\}$ be the Bayes probability of error for testing the simple hypotheses (2.1). If $0 < \mathbb{P}\{H_0\} < 1$ then*

$$\inf_S \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_n^B = -\Psi^\sharp(0).$$

The infimum is taken over all tests for (2.1) with error probabilities given by (2.2).

Moreover,

$$\Psi^\sharp(0) = -K_L(g_{\alpha_0}; f),$$

with $K_L(f; g)$ given above and

$$g_{\alpha_0} = \frac{fg}{\alpha_0 f + (1 - \alpha_0)g}.$$

The value of α_0 is determined in the proof of this result.

Remark 2.3. *Hirukawa and Taniguchi [30] have studied non-Gaussian locally stationary processes, focusing on proving local asymptotic normality (LAN). We comment on possible relations to our results at subsection 2.2.1.*

This chapter is organized as follows. In Section 2.2 we derive a large deviation result for the log-likelihood ratio for testing simple hypotheses for locally stationary Gaussian processes. Also in this section, we present a study about the rates of exponential decay of the type I and type II error probabilities given in (2.2), as a result we obtain a Hoeffding bound for testing the simple hypotheses (2.1). A proof of Proposition 2.1 is given in

Section 2.3 along with an introduction that refers to the classical Stein's lemma for i.i.d. observations. In the same spirit, Section 2.4 presents the classical Chernoff bound and the corresponding proof of Proposition 2.2. Finally, some technical details used in our proofs are relegated to the Appendix.

2.2 Large deviations for locally stationary Gaussian processes

We begin this section with some definitions. Let \mathbf{P}_n and \mathbf{Q}_n be probability measures defined on some space $(\Omega_n, \mathcal{F}_n)$. For each n , suppose that \mathbf{Q}_n is absolutely continuous w.r.t. \mathbf{P}_n (in short notation $\mathbf{Q}_n \ll \mathbf{P}_n$). Let $\Lambda_n = \log \frac{d\mathbf{Q}_n}{d\mathbf{P}_n}$ be the log-likelihood ratio between \mathbf{Q}_n and \mathbf{P}_n . Let μ_n be a measure dominating \mathbf{P}_n and \mathbf{Q}_n . Since $d\mathbf{P}_n = \mathbf{p}_n d\mu_n$ and $d\mathbf{Q}_n = \mathbf{q}_n d\mu_n$, where \mathbf{p}_n and \mathbf{q}_n are the Radon-Nykodim derivatives of \mathbf{P}_n and \mathbf{Q}_n w.r.t. μ_n , $\Lambda_n = \log \frac{\mathbf{q}_n}{\mathbf{p}_n}$.

Under \mathbf{P}_n , let $\Phi_n(\cdot)$ be the m.g.f. of Λ_n ; $\Phi_n(\alpha) = \mathbb{E}_{\mathbf{P}_n}[\exp\{\alpha \Lambda_n\}]$. Note that for any $\alpha \in [0, 1]$

$$\Phi_n(\alpha) = \int \mathbf{p}_n^{1-\alpha}(x) \mathbf{q}_n^\alpha(x) d\mu_n(x). \quad (2.3)$$

The RHS of (2.3) is known as the Hellinger transform between \mathbf{P}_n and \mathbf{Q}_n . For zero mean Gaussian processes, i.e., $\mathbf{P}_n \sim \mathcal{N}_n(0, \Gamma_n)$, $\mathbf{Q}_n \sim \mathcal{N}_n(0, \widetilde{\Gamma}_n)$, Taniguchi and Kakizawa [46] have introduced the terms *asymptotic Kullback-Leibler divergence* and *asymptotic Chernoff information with index α* , $0 < \alpha < 1$, as

$$K_L = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_{\mathbf{P}_n} \left[\log \frac{d\mathbf{P}_n}{d\mathbf{Q}_n} \right] \quad (2.4)$$

and

$$C_\alpha = \lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{E}_{\mathbf{P}_n} \left[\left\{ \frac{d\mathbf{Q}_n}{d\mathbf{P}_n} \right\}^\alpha \right], \quad (2.5)$$

respectively. See Section 7.6 in [46].

Following these definitions we have that for any two equivalent Gaussian measures \mathbf{P}_n and \mathbf{Q}_n

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \Phi_n(\alpha) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \int \mathbf{p}_n^{1-\alpha}(x) \mathbf{q}_n^\alpha(x) dx = -C_\alpha. \quad (2.6)$$

That is, computing the asymptotic Chernoff divergence is equivalent to knowing the logarithmic limit of the Hellinger transform which in turn, and thanks to Bouaziz' lemma below, is equivalent to establishing a large deviation principle for the log-likelihood ratio of \mathbf{Q}_n w.r.t \mathbf{P}_n . This is summarized by the equation (2.6).

Bouaziz [8] showed the following general result on large deviations. Let S_n be a sequence of real-valued random variables defined on some probability space $(\Omega_n^*, \mathcal{F}_n^*, \mathbb{P}_n)$. Let ϕ_n be the moment generating function of S_n , i.e., $\phi_n(\alpha) = \mathbb{E}_{\mathbb{P}_n}[\exp\{\alpha S_n\}]$.

Bouaziz' Lemma (Lemma 2.1 in [8]). *Suppose that*

1. $\phi_n(1) < \infty$ for all (sufficiently large) n ,
2. $(1/n) \log \phi_n(\alpha) \rightarrow \psi(\alpha) \forall \alpha \in [0, 1]$,
3. ψ is differentiable and strictly convex on $(0, 1)$.

Then for all $a \in (\psi'(0), \psi'(1))$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_n\{S_n > na\} = -\psi^\sharp(a),$$

where ψ^\sharp is the Fenchel-Legendre conjugate of ψ defined by

$$\psi^\sharp(a) = \sup_{0 \leq \alpha \leq 1} [a\alpha - \psi(\alpha)].$$

In addition, the same result holds for $\{S_n \geq na\}$.

We will apply this lemma to the log-likelihood Λ_n of two locally stationary Gaussian processes. More precisely, let $X = (X_{k,n})_{1 \leq k \leq n}$ and $\tilde{X} = (\tilde{X}_{k,n})_{1 \leq k \leq n}$ be realizations of

locally stationary Gaussian processes with probability laws \mathbf{P}_n , \mathbf{Q}_n and time-varying spectral densities f , g , respectively. From now on $\mathbf{P}_n \sim \mathcal{N}_n(0, \Sigma_n)$ and $\mathbf{Q}_n \sim \mathcal{N}_n(0, \widetilde{\Sigma}_n)$ where Σ_n and $\widetilde{\Sigma}_n$ are defined in (1.5), cf. Ch. 1. Let $\Lambda_n = \log \frac{d\mathbf{Q}_n}{d\mathbf{P}_n}$ be the log-likelihood ratio of \mathbf{Q}_n w.r.t. \mathbf{P}_n and $\Phi_n(\cdot)$ its moment generating function. From (2.3), it can be seen that $\Phi_n(1) = 1$ for all n . Let \mathbb{T} and $\mathcal{B}_{\mathbb{T}}$ be as in the introduction of this chapter. We have the following

Lemma 2.4. *If $f, g \in \mathcal{B}_{\mathbb{T}}$ then for any $\alpha \in [0, 1]$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \Phi_n(\alpha) = \Psi(\alpha),$$

where

$$\Psi(\alpha) = \frac{1}{4\pi} \int_{\mathbb{T}} \left[\alpha \log \frac{f}{g} - \log \left(\frac{\alpha f + (1 - \alpha)g}{g} \right) \right] d\lambda du. \quad (2.7)$$

Proof. Since $\Phi_n(1) = \Phi_n(0) = 1$, the result is trivial for $\alpha \in \{0, 1\}$. For $\alpha \in (0, 1)$ we show in the Appendix that for any $n \in \mathbb{N}$ we have that

$$\begin{aligned} \frac{1}{n} \log \Phi_n(\alpha) &= \frac{\alpha}{2n} (\log \det \Sigma_n - \log \det \widetilde{\Sigma}_n) \\ &\quad - \frac{1}{2n} (\log \det [\alpha \Sigma_n + (1 - \alpha) \widetilde{\Sigma}_n] - \log \det \widetilde{\Sigma}_n). \end{aligned}$$

Since $f, g \in \mathcal{B}_{\mathbb{T}}$ it follows that for any $\alpha \in (0, 1)$, $\alpha f + (1 - \alpha)g \in \mathcal{B}_{\mathbb{T}}$. An application of Theorem 3.2(ii) of Dahlhaus [18] yields

$$\begin{aligned} \frac{1}{n} \log \det \Sigma_n &\rightarrow \frac{1}{2\pi} \int_{\mathbb{T}} \log 2\pi f d\lambda du, \\ \frac{1}{n} \log \det \widetilde{\Sigma}_n &\rightarrow \frac{1}{2\pi} \int_{\mathbb{T}} \log 2\pi g d\lambda du, \\ \frac{1}{n} \log \det [\alpha \Sigma_n + (1 - \alpha) \widetilde{\Sigma}_n] &\rightarrow \frac{1}{2\pi} \int_{\mathbb{T}} \log 2\pi (\alpha f + (1 - \alpha)g) d\lambda du. \end{aligned}$$

The result now follows by summing up all the terms above. □

According to (2.5) and (2.7), the asymptotic Chernoff information between two locally stationary Gaussian processes is given by the integral

$$C_\alpha = \frac{1}{4\pi} \int_{\mathbb{T}} \left[\log \left(\frac{\alpha f + (1-\alpha)g}{g} \right) - \alpha \log \frac{f}{g} \right] d\lambda du.$$

In the Appendix we show that (2.4) is given by the quantity

$$K_L = \frac{1}{4\pi} \int_{\mathbb{T}} \left(\log \frac{g}{f} + \frac{f-g}{g} \right) d\lambda du.$$

In accordance with the terminology introduced by Taniguchi and Kakizawa, this expression is called the asymptotic Kullback-Leibler divergence for locally stationary Gaussian processes.

Observe that when $f(u, \lambda)$ and $g(u, \lambda)$ do not depend on u the two expressions above generalize well-known formulas for the asymptotic Chernoff information between stationary Gaussian sequences (Coursol and Dacunha-Castelle [13]) and for the asymptotic Kullback-Leibler divergence between stationary Gaussian sequences (obtained by Taniguchi [45] as the approximate slope of the primitive likelihood ratio test between stationary Gaussian processes with spectral density $f(\lambda)$).

Let f, g and Ψ be as in Lemma 2.4. Since the function $h(x) = -\log(1+\theta x)$ is strictly convex $\forall \theta \neq 0$ it follows that $\Psi(\cdot)$ is strictly convex whenever $f \neq g$. In addition, the function $\Psi(\cdot)$ is differentiable and $\Psi'(0, 1) = (\Psi'(0), \Psi'(1))$ where

$$\Psi'(0) = \frac{1}{4\pi} \int_{\mathbb{T}} \left[\log \frac{f}{g} - \frac{f-g}{g} \right] d\lambda du \quad (2.8)$$

$$\Psi'(1) = \frac{1}{4\pi} \int_{\mathbb{T}} \left[\log \frac{f}{g} - \frac{f-g}{f} \right] d\lambda du. \quad (2.9)$$

Since $\log(x) < x - 1$, $\forall x > 0$, $x \neq 1$ it follows that $\Psi'(0) < 0 < \Psi'(1)$ whenever $f \neq g$. So, we have found conditions under which the m.g.f., $\Phi_n(\cdot)$, of the log-likelihood ratio between two locally stationary Gaussian processes satisfies the assumptions of Bouaziz' lemma. We have the following

Theorem 2.5. *Let \mathbf{P}_n and \mathbf{Q}_n be probability measures corresponding to two locally stationary Gaussian processes with time-varying spectral densities f and g , respectively. Let $f, g \in \mathcal{B}_{\mathbb{T}}$ and suppose additionally that $f \neq g$ on \mathbb{T} . Let Λ_n be the log-likelihood ratio of \mathbf{Q}_n w.r.t. \mathbf{P}_n . Then for all $a \in (\Psi'(0), \Psi'(1))$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}_n \{ \Lambda_n > na \} = -\Psi^\#(a),$$

where

$$\Psi^\#(a) = \sup_{0 \leq \alpha \leq 1} [a\alpha - \Psi(\alpha)],$$

with $\Psi(\alpha)$ given as in (2.7). The same is true for $\{ \Lambda_n \geq na \}$.

This result is a large deviation for the log-likelihood ratio between the laws of two locally stationary Gaussian processes. Large deviations for quadratic forms of locally stationary processes have been obtained by Zani [49]. We now use Theorem (2.5) to establish exponential rates of decrease for the false-alarm and nondetection rates for testing the hypotheses considered in (2.1). Following Bouaziz let us recall that given a false-alarm level α_n in $(0, 1)$, the most powerful tests of \mathbf{P}_n against \mathbf{Q}_n at the level α_n are the Neyman-Pearson tests of size α_n , i.e. those tests τ_n which satisfy

$$\mathbb{I}_{\{\Lambda_n > a_n\}} \leq \tau_n \leq \mathbb{I}_{\{\Lambda_n \geq a_n\}}, \quad \mathbf{E}_{\mathbf{P}_n}[\tau_n] = \alpha_n, \quad \mathbf{E}_{\mathbf{Q}_n}[\tau_n] = 1 - \beta_n.$$

For every $n \in \mathbb{N}$ we have the following inequalities:

$$\frac{1}{n} \log \mathbf{P}_n \{ \Lambda_n > a_n \} \leq \frac{1}{n} \log \alpha_n \leq \frac{1}{n} \log \mathbf{P}_n \{ \Lambda_n \geq a_n \}.$$

Under the conditions of Theorem 2.5 and assuming that $a_n \sim na$, it follows that for all $a \in (\Psi'(0), \Psi'(1))$, $\frac{1}{n} \log \alpha_n(a) \rightarrow -\Psi^\#(a)$. Let $\widehat{\Phi}_n(\alpha) = \mathbf{E}_{\mathbf{Q}_n}[\exp\{\alpha \widetilde{\Lambda}_n\}]$ be the m.g.f. of $\widetilde{\Lambda}_n = \log \frac{d\mathbf{P}_n}{d\mathbf{Q}_n}$ under \mathbf{Q}_n . Observe that $\widetilde{\Lambda}_n = -\Lambda_n$. A straightforward calculation allows us to see that $\widehat{\Phi}_n(\alpha) = \Phi_n(1 - \alpha)$, for all $\alpha \in [0, 1]$. Consequently,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \widehat{\Phi}_n(\alpha) \rightarrow \Psi(1 - \alpha).$$

The Fenchel-Legendre conjugate of this limiting function is given by

$$\widehat{\Psi}^\sharp(a) = \sup_{\alpha \in [0,1]} [a\alpha - \Psi(1 - \alpha)] = a + \Psi^\sharp(-a).$$

Thus, under conditions of Theorem 1 we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{Q}_n \{ \widetilde{\Lambda}_n > na \} = -a - \Psi^\sharp(-a),$$

and the same holds for $\{\widetilde{\Lambda}_n \geq na\}$.

Since $\{\Lambda_n < nb\} = \{\widetilde{\Lambda}_n > -nb\}$ it follows that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{Q}_n \{ \Lambda_n < nb \} = b - \Psi^\sharp(b).$$

Thus, for Neyman-Pearson tests and for every $n \in \mathbb{N}$ we have the following inequalities:

$$\frac{1}{n} \log \mathbf{Q}_n \{ \Lambda_n < a_n \} \leq \frac{1}{n} \log \beta_n \leq \frac{1}{n} \log \mathbf{Q}_n \{ \Lambda_n \leq a_n \}.$$

Hence for locally stationary Gaussian processes satisfying the conditions of Theorem 1 we conclude that for all $a \in (\Psi'(0), \Psi'(1))$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n(a) = a - \Psi^\sharp(a),$$

where $\Psi^\sharp(\cdot)$ is the Fenchel-Legendre conjugate of $\Psi(\cdot)$ given in (2.7). Summarizing, we have the following

Corollary 2.6. *[Hoeffding bound] Under the assumptions of Theorem 1. Let $\Psi'(0)$ and $\Psi'(1)$ be given as in (2.8) and (2.9), respectively. Then for all $a \in (\Psi'(0), \Psi'(1))$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \alpha_n(a) = -\Psi^\sharp(a) < 0, \tag{2.10}$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n(a) = a - \Psi^\sharp(a) < 0, \tag{2.11}$$

where

$$\Psi^\sharp(a) = \frac{1}{4\pi} \int_{\mathbb{T}} \left[\log \left(\frac{\alpha_a f + (1 - \alpha_a)g}{g} \right) + \frac{\alpha_a(g - f)}{\alpha_a f + (1 - \alpha_a)g} \right] d\lambda du,$$

with α_a the unique solution to $\Psi'(\alpha_a) = a$.

Proof. It remains to establish the expression for $\Psi^\sharp(a)$. Define $G(a, \alpha) = a\alpha - \Psi(\alpha)$. Observe that $\frac{\partial}{\partial \alpha} G(a, \alpha) = a - \Psi'(\alpha)$, $\frac{\partial^2}{\partial^2 \alpha} G(a, \alpha) = -\Psi''(\alpha)$. Since for all $\alpha \in [0, 1]$, $\Psi''(\alpha) = (4\pi)^{-1} \int_{\mathbb{T}} (f - g)^2 / [\alpha f + (1 - \alpha)g]^2 d\lambda du > 0$, the function $\frac{\partial}{\partial \alpha} G(a, \alpha)$ is strictly decreasing on $(\Psi'(0), \Psi'(1))$. Then, the equation $a = \Psi'(\alpha)$ has a unique solution α_a , i.e. there exists a unique $\alpha_a \in [0, 1]$ such that $a = \Psi'(\alpha_a)$. It follows that $\Psi^\sharp(a) = G(a, \alpha_a)$. A straightforward calculation yields

$$\begin{aligned} G(a, \alpha_a) &= \frac{1}{4\pi} \int_{\mathbb{T}} \left[\log \left(\frac{\alpha_a f + (1 - \alpha_a)g}{g} \right) + \frac{\alpha_a(g - f)}{\alpha_a f + (1 - \alpha_a)g} \right] d\lambda du \\ &= K_L(g_{\alpha_a}, f), \end{aligned} \tag{2.12}$$

where $K_L(f, g)$ is given in Proposition 2.1 and

$$g_{\alpha_a} = \frac{fg}{\alpha_a f + (1 - \alpha_a)g}. \quad \square$$

Results of this type can be found in the literature on large deviations for stochastic processes. For instance, Dembo and Zeitouni [28] (Theorem 3.4.3 §4.2) have obtained Corollary 1 for the case of a sequence of n i.i.d. random variables, Theorem 2.4 in [8] and Lemma 8.2.3 in [46] treat the same problem for stationary Gaussian sequences.

In the information theoretic literature this type of result is referred to as the Hoeffding-Blahut-Csiszár-Longo bound. Recently, Audenaert et al. [1] have obtained a Hoeffding bound in quantum hypothesis testing, and Gapeev and K  chler [27] have provided similar large deviation results for testing Ornstein-Uhlenbeck-type models.

An interpretation of Corollary 2.6 is that for every $a \in (\Psi'(0), \Psi'(1))$ there exists a test τ_a whose type I and type II error probabilities, $\alpha_n(a)$ and $\beta_n(a)$, satisfy (2.10) and (2.11), respectively. Moreover, the ratio $\frac{\alpha_n(a)}{\beta_n(a)}$ behaves roughly like e^{-na} , and we have the following cases:

1. For $a > 0$, $\frac{\alpha_n(a)}{\beta_n(a)} \rightarrow 0$ as $n \rightarrow \infty$. That is, $\alpha_n(a) \rightarrow 0$ and $\beta_n(a) \rightarrow c_0$.

2. For $a = 0$, $\frac{\alpha_n(a)}{\beta_n(a)} \rightarrow 1$ as $n \rightarrow \infty$. That is, for n large enough $\alpha_n(a)$ and $\beta_n(a)$ share the same growth rate.
3. For $a < 0$, $\frac{\alpha_n(a)}{\beta_n(a)} \rightarrow \infty$ as $n \rightarrow \infty$. That is, $\alpha_n(a) \rightarrow c_1$ and $\beta_n(a) \rightarrow 0$.

In the next sections we study these three cases in more detail. For instance, the rate with which $\beta_n \rightarrow 0$, case $a < 0$, is given by Stein's Lemma and the rate in which both type I error and type II error grow equally, case $a = 0$, appears as the Chernoff bound for testing simple hypotheses.

2.2.1 A comment on large deviations for non-Gaussian locally stationary processes

To our knowledge, Hirukawa and Taniguchi [30] have presented the only systematic study on non-Gaussian locally stationary processes. We now present an outline of a possible extension of our large deviation results to the non-Gaussian processes introduced in [30].

Let us begin by saying that a non-Gaussian locally stationary process satisfies the representation given in (1.1):

$$X_{k,n} = \int_{-\pi}^{\pi} \exp(i\lambda k) A_{k,n}^{\circ}(\lambda) d\epsilon(\lambda).$$

As in Definition 1.1, ϵ is a complex-valued stochastic process defined on $[-\pi, \pi]$ but it is not necessarily a Brownian motion. Instead, it is supposed that

$$\text{cum} \{ d\epsilon(\lambda_1), \dots, d\epsilon(\lambda_k) \} = \eta \left(\sum_{i=1}^k \lambda_i \right) h_k(\lambda_1, \dots, \lambda_{k-1}) d\lambda_1 \cdots d\lambda_k,$$

where cum stands for the cumulant of k -th order, η is a 2π -periodic extension of the Dirac delta function and $|h_k(\lambda_1, \dots, \lambda_{k-1})| \leq C_k$. As in Definition 1.2 the trend function $A_{k,n}^{\circ}(\cdot)$

has a uniformly approximating function $A(k/n, \cdot)$ which is assumed to be uniformly bounded from above and bounded away from zero as well as continuous in the first component.

Since Hirukawa and Taniguchi [30] were focused on locally asymptotic normality (LAN), it was natural to assume that the process had a parametric structure. Namely, realizations $(X_{k,n})_{1 \leq k \leq n}$ of a non-Gaussian locally stationary process with trend function A_θ° , approximating function A_θ and time-varying function $f_\theta(u, \lambda) := |A_\theta(u, \lambda)|^2$ where $\theta = (\theta_1, \dots, \theta_q) \in \Theta \in \mathbb{R}^q$ have been considered.

In [30] the following assumptions were made:

A1. $A_\theta(u, \lambda)$, the gradient $\nabla A_\theta(u, \lambda)$, and the Hessian $\nabla^2 A_\theta(u, \lambda)$ have components which are differentiable in u and λ with uniformly continuous derivative $\frac{\partial}{\partial u} \frac{\partial}{\partial \lambda}$. Also, it is assumed that the gradient and Hessian of $A_{\theta,k,n}^\circ(\lambda) - A_\theta(k/n, \lambda)$ are uniformly bounded in k and λ .

A2. Write

$$\xi_k = \int_{-\pi}^{\pi} \exp(ik\lambda) d\epsilon(\lambda),$$

and assume that the ξ_k 's are i.i.d. random variable with $E[\xi_k] = 0$, $E[\xi_k^2] = 1$ and $E[\xi_k^4] < \infty$. Furthermore, assume that the distribution of ξ_k is absolutely continuous w.r.t. Lebesgue measure and that the corresponding probability density p satisfies regularity conditions such as being light-tailed and having smooth first two derivatives p' and p'' .

A3. Assume that $\{X_{k,n}\}$ has the MA(∞) and AR(∞) representations

$$X_{k,n} = \sum_{j=0}^{\infty} a_{\theta,k,n}^\circ(j) \xi_{k-j},$$

$$a_{\theta,k,n}^\circ(0) \xi_k = \sum_{t=0}^{\infty} b_{\theta,k,n}^\circ(t) X_{k-t,n}.$$

It is also assumed that the non-random coefficients $a_{\theta,k,n}^\circ(\cdot)$ and $b_{\theta,k,n}^\circ(\cdot)$ satisfy some smoothness and summability conditions.

Let $\mathbf{P}_{\theta,n}$ be the probability distribution of $(\xi_s, s \leq 0, X_{1,n}, \dots, X_{n,n})$. Then under assumptions A1, A2 and A3, the log-likelihood ratio between two non-Gaussian locally stationary processes defined via the points $\theta, \theta_n \in \Theta$ is given by

$$\Lambda_n(\theta, \theta_n) = \log \frac{d\mathbf{P}_{\theta_n,n}}{d\mathbf{P}_{\theta,n}} = 2 \sum_{k=1}^n \log \Phi_{k,n}(\theta, \theta_n),$$

where

$$\Phi_{k,n}^2(\theta, \theta_n) = \frac{g_{\theta_n,k,n}(z_{\theta,k,n} + q_{k,n})}{g_{\theta,k,n}(z_{\theta,k,n})}$$

with

$$\begin{aligned} g_{\theta,k,n}(\cdot) &= \frac{1}{a_{\theta,k,n}^\circ(0)} p\left(\frac{\cdot}{a_{\theta,k,n}^\circ(0)}\right), \quad z_{\theta,k,n} = a_{\theta,k,n}^\circ(0) \xi_k \\ a_{\theta,k,n}^\circ(0) &\equiv \exp\left\{\frac{1}{4\pi} \int_{-\pi}^{\pi} \log f_{\theta,k,n}(\lambda) d\lambda\right\}, \quad f_{\theta,k,n}(\lambda) = |A_{\theta,k,n}^\circ(\lambda)|^2 \\ q_{k,n} &= \sum_{t=1}^{k-1} \left[\frac{1}{\sqrt{n}} h^\top \nabla b_{\theta,k,n}^\circ(t) + \frac{1}{2n} h^\top \nabla^2 b_{\theta,k,n}^\circ(t) h \right] X_{k-t,n} \\ &\quad + \sum_{r=0}^{\infty} \frac{1}{\sqrt{n}} h^\top \nabla c_{\theta^{**},k,n}^\circ(r) \xi_{-r} \\ c_{\theta,k,n}^\circ(r) &= \sum_{s=0}^r b_{\theta,k,n}^\circ(k+s) a_{\theta,0,n}^\circ(r-s). \end{aligned}$$

The points θ^* and θ^{**} belong to the segment defined by θ and $\theta + \frac{h}{\sqrt{n}}$, for some $h \in \mathbb{R}^q$.

As shown in the present paper, to establish a large deviation for the log-likelihood ratio between non-Gaussian locally stationary processes we could start by studying the behavior of

$$\frac{1}{n} \log \mathbb{E}_{\mathbf{P}_{\theta,n}} [\exp \{s \Lambda_n(\theta, \theta_n)\}]$$

as $n \rightarrow \infty$ with $\Lambda_n(\theta, \theta_n)$ as given above. This task is beyond the scope of the present work; however, as it will be illustrated in the next sections such a large deviation re-

sult might prove useful to determine the asymptotic efficiency of Neyman-Pearson and Bayes tests even in a non-Gaussian setting.

2.3 Proof of Stein's lemma

In the i.i.d. case, Stein's lemma for testing simple hypotheses can be stated as follows. Let $X = (X_1, \dots, X_n)$ be a sequence of i.i.d. random variables. Consider the following hypotheses

$$H_0 : X \sim p^n \quad \text{vs} \quad H_1 : X \sim q^n. \quad (2.13)$$

The Neyman-Pearson lemma dictates to use the log-likelihood ratio test between q^n and p^n to minimize the type II error probability, β_n , of this test. Stein's lemma tells us that, when the type I error probability, α_n , of this test is bounded then β_n will converge to zero at an exponential rate given by the Kullback-Leibler divergence between p and q . More precisely, and following Dembo and Zeitouni [24], we have

Stein's Lemma. *Let β_n^ϵ be the infimum of β_n among all tests whose false-alarm rate $\alpha_n < \epsilon$. Then for any $\epsilon < 1$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^\epsilon = -K_L(p; q),$$

where $K_L(p; q)$ is the Kullback-Leibler information divergence between p and q .

Many authors have provided proofs of this lemma. For instance, Lemma 6.1 in Bahadur [2], Section B, Chapter 6 in Bucklew [10], Lemma 3.4.6 in [24] or the original statement, Theorem 2, in Chernoff [12] all present proofs of the so-called Stein's lemma. Roughly speaking, in an i.i.d. framework when testing between two probability

measures, p and q , if the false-alarm detection rate of the test statistic remains bounded when $n \rightarrow \infty$ then the nondetection rate, β_n , will decrease to zero exponentially fast.

We have stated an analogue of Stein's lemma for locally stationary Gaussian processes in the introduction of this chapter. In this section we present a proof of this lemma which is based on Corollary 2.6. From now on we assume that $a_n \sim na$, that \mathbf{Q}_n is absolutely continuous w.r.t. \mathbf{P}_n and that $\Psi'(0) = -K_L(f, g)$ where $\Psi'(0)$ is given in (2.8).

We now give a proof of Proposition 2.1.

Proof of Proposition 2.1. Let $\varepsilon > 0$ be given. Let β_n^ε be the infimum of β_n among all tests with $\alpha_n < \varepsilon$. It is enough to consider Neyman-Pearson tests, i.e., those tests τ_n such that

$$\mathbb{I}_{\{\Lambda_n > a_n\}} \leq \tau_n \leq \mathbb{I}_{\{\Lambda_n \geq a_n\}}, \quad \mathbb{E}_{\mathbf{P}_n}[\tau_n] = \alpha_n, \quad \mathbb{E}_{\mathbf{Q}_n}[\tau_n] = 1 - \beta_n. \quad (2.14)$$

We get that

$$\frac{1}{n} \log \mathbf{Q}_n\{\Lambda_n < a_n\} \leq \frac{1}{n} \log \beta_n \leq \frac{1}{n} \log \mathbf{Q}_n\{\Lambda_n \leq a_n\}. \quad (2.15)$$

Since $\mathbf{Q}_n \ll \mathbf{P}_n$ we have $\int \mathbb{I}_{\{\Lambda_n \leq a_n\}} \mathbf{e}^{\Lambda_n} d\mathbf{P}_n = \int \mathbb{I}_{\{\Lambda_n \leq a_n\}} d\mathbf{Q}_n$. This implies that

$$\frac{1}{n} \log \mathbf{Q}_n\{\Lambda_n \leq a_n\} \leq a. \quad (2.16)$$

For any $a \in (\Psi'(0), \Psi'(1))$ there exists a test, τ_a say, such that its error probabilities of first type, $\alpha_n(a)$, and second type, $\beta_n(a)$, satisfy (2.10) and (2.11). For n large enough, $\alpha_n(a) < \varepsilon$. Hence $\frac{1}{n} \log \beta_n^\varepsilon < \frac{1}{n} \log \beta_n(a)$. Now, use the RHS of (2.15) and (2.16) to deduce that $\frac{1}{n} \log \beta_n^\varepsilon \leq \Psi'(0)$. Therefore $\forall \delta > 0$ we get that

$$\limsup_n \frac{1}{n} \log \beta_n^\varepsilon \leq \Psi'(0) + \delta.$$

To continue with the proof we need the following law of large numbers for Λ_n . For every $\theta \in (0, 1)$, define $d\mathbf{P}_{n,\theta} = [\mathbf{e}^{\theta\Lambda_n} / \Psi_n(\theta)] d\mathbf{P}_n$, $\Psi_n = \frac{1}{n} \log \Phi_n$. Then

$$\frac{1}{n} \Lambda_n \rightarrow \Psi'(\theta) \text{ in } \mathbf{P}_{n,\theta} \text{ probability } 1. \quad (2.17)$$

The argument to show that (2.17) holds is contained in the proof of Lemma 2.1 in [8].

Choose $a \in (\Psi'(0), \Psi'(1))$ whose corresponding test, τ_a , is such that for n large enough $\alpha_n(a) = \mathbf{P}_n\{\Lambda_n > a_n\} \leq \varepsilon$. The law of large numbers for Λ_n just presented implies that for all $\delta > 0$

$$\liminf_n \mathbf{P}_n\{\Lambda_n \in [n(\Psi'(0) - \delta), a_n]\} \geq 1 - \varepsilon. \quad (2.18)$$

Use the LHS of the inequality in (2.14) to get that

$$\begin{aligned} \beta_n &\geq \int \mathbb{1}_{\{\Lambda_n < a_n\}} d\mathbf{Q}_n = \int \mathbb{1}_{\{\Lambda_n < a_n\}} \mathbf{e}^{\Lambda_n} d\mathbf{P}_n \\ &\geq \int \mathbb{1}_{\{\Lambda_n \in [n(\Psi'(0) - \delta), a_n]\}} \mathbf{e}^{\Lambda_n} d\mathbf{P}_n \\ &\geq \mathbf{e}^{n(\Psi'(0) - \delta)} \int \mathbb{1}_{\{\Lambda_n \in [n(\Psi'(0) - \delta), a_n]\}} d\mathbf{P}_n. \end{aligned}$$

Hence for all $\delta > 0$

$$\frac{1}{n} \log \beta_n(a) \geq \Psi'(0) - \delta + \frac{1}{n} \log \mathbf{P}_n\{\Lambda_n \in [n(\Psi'(0) - \delta), a_n]\}.$$

Finally, use (2.18) and recall that by assumption $\varepsilon < 1$ to deduce that for all $\delta > 0$

$$\liminf_n \frac{1}{n} \log \beta_n^\varepsilon \geq \Psi'(0) - \delta.$$

This completes the proof. □

We have extended Stein's Lemma from the classical i.i.d setup to nonstationary processes. In the context of model selection, Dahlhaus [18] found $\Psi'(0)$ and termed it

as the Kullback-Leibler divergence between two locally stationary processes with time-varying spectral densities f and g . According to the conclusion of Proposition 2.1 we can now adopt the notation $K_L(f, g)$ and refer to it as the asymptotic Kullback-Leibler divergence between two locally stationary Gaussian processes.

2.4 Proof of Chernoff bound

Consider the setting given in (2.13). Because of the Neyman-Pearson lemma, to study the asymptotic efficiency of any test in (2.13) it is sufficient to have some *a priori* probability on H_0 and to consider the two types of error probabilities given by

$$\alpha_n = \mathbb{P}\{S_n \geq a_n \mid H_0\}, \quad \beta_n = \mathbb{P}\{S_n < a_n \mid H_1\},$$

where $S_n = \sum_{k=1}^n X_k$. A principle to find the threshold a_n is that of minimizing $\alpha_n + \lambda \beta_n$ for some $\lambda > 0$. Basically, this criterion is equivalent to requiring that as $n \rightarrow \infty$, α_n and β_n approach zero at the same rate. A solution to this problem was provided by Chernoff [11]. Namely, we have

Chernoff Bound.

$$\lim_{n \rightarrow \infty} [\inf_{a_n} (\beta_n + \lambda \alpha_n)]^{1/n} = \inf_{0 < t < 1} \int p(x)^{1-t} q(x)^t dx.$$

Dembo and Zeitouni [24] consider the same problem. Their solution is mainly based on elements of large deviations for the log-likelihood ratio test. Taniguchi and Kakizawa [46] have obtained the Chernoff bound for testing simple hypotheses for stationary processes. Their solution is also mainly based on large deviations. The content of Proposition 2.2 is an extension of these results to the framework of locally stationary Gaussian processes. We now present a proof of such proposition.

Proof of Proposition 2.2. Again it is sufficient to consider Neyman-Pearson tests, τ_n . Let $\alpha_n(0)$ and $\beta_n(0)$ be the error probabilities of the Neyman-Pearson test with zero threshold. For any other Neyman-Pearson test, τ_n , either $\alpha_n \geq \alpha_n(0)$ ($a_n \leq 0$) or $\beta_n(0) \leq \beta$ ($a_n \geq 0$). Thus the inequalities

$$\begin{aligned} \min\{\alpha_n(0), \beta_n(0)\} \min\{\mathbb{P}\{H_0\}, \mathbb{P}\{H_1\}\} &\leq \mathbb{P}_n^B \\ &\leq 2 \max\{\alpha_n(0), \beta_n(0)\} \max\{\mathbb{P}\{H_0\}, \mathbb{P}\{H_1\}\} \end{aligned}$$

yield

$$\inf_S \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_n^B \geq \liminf_{n \rightarrow \infty} \min\left\{\frac{1}{n} \log \alpha_n(0), \frac{1}{n} \log \beta_n(0)\right\}$$

and

$$\inf_S \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_n^B \leq \limsup_{n \rightarrow \infty} \max\left\{\frac{1}{n} \log \alpha_n(0), \frac{1}{n} \log \beta_n(0)\right\}.$$

Recall that $0 < \mathbb{P}\{H_0\} < 1$ and S denotes the set of all tests for (2.1) with probability errors given by (2.2). Given (2.10) and (2.11)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \alpha_n(0) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n(0) = -\Psi^\#(0).$$

Observe that by definition

$$-\Psi^\#(0) = -\sup_{\alpha \in [0,1]} [-\Psi(\alpha)] = \inf_{\alpha \in [0,1]} \Psi(\alpha) = f(0, t_0)$$

where $f(0, t_0)$ is given by (2.12). This completes the proof. \square

2.5 Appendix

In this section we use the notation introduced in Section 2.2 and the assumptions of Lemma 2.4. Consider the log-likelihood of \mathbf{Q}_n w.r.t. \mathbf{P}_n , $\Lambda_n = \log \frac{d\mathbf{Q}_n}{d\mathbf{P}_n}$, and its moment generating function with respect to \mathbf{P}_n , i.e., the function $\Phi_n : [0, 1] \rightarrow [0, 1]$ given by

$$\Phi_n(\alpha) = \mathbb{E}_{\mathbf{P}_n} [\exp \{\alpha \Lambda_n\}] = \mathbb{E}_{\mathbf{P}_n} \left\{ \frac{d\mathbf{Q}_n}{d\mathbf{P}_n} \right\}^\alpha = \int \mathbf{p}_n^{1-\alpha}(x) \mathbf{q}_n^\alpha(x) dx.$$

Let us calculate this integral:

$$\begin{aligned}
\Phi_n(\alpha) &= \int \det(2\pi\Sigma_n)^{-(1-\alpha)/2} \mathbf{e}^{-\frac{(1-\alpha)}{2}x^\top \Sigma_n^{-1}x} \det(2\pi\widetilde{\Sigma}_n)^{-(\alpha)/2} \mathbf{e}^{-\frac{\alpha}{2}x^\top \widetilde{\Sigma}_n^{-1}x} dx \\
&= \det(\Sigma_n)^{-(1-\alpha)/2} \det(\widetilde{\Sigma}_n)^{-\alpha/2} \det[(1-\alpha)\Sigma_n^{-1} + \alpha\widetilde{\Sigma}_n^{-1}]^{-1/2} \\
&= \det(\Sigma_n)^{\alpha/2} \det(\widetilde{\Sigma}_n)^{(1-\alpha)/2} \det[\alpha\Sigma_n + (1-\alpha)\widetilde{\Sigma}_n]^{-1/2}.
\end{aligned}$$

The latter follows after multiplying the integrand by a proper constant, using the fact that every probability density function integrates to 1 and using properties of the determinant.

Therefore

$$\log \Phi_n(\alpha) = \frac{1}{2} \left(\alpha \log \frac{\det \Sigma_n}{\det \widetilde{\Sigma}_n} - \log \frac{\det[\alpha\Sigma_n + (1-\alpha)\widetilde{\Sigma}_n]}{\det \widetilde{\Sigma}_n} \right).$$

Similar calculations show that the m.g.f. of the log-likelihood ratio of \mathbf{P}_n w.r.t. \mathbf{Q}_n is equal to

$$\log \frac{d\mathbf{P}_n}{d\mathbf{Q}_n}(x) = -\frac{1}{2} \log \frac{\det \Sigma_n}{\det \widetilde{\Sigma}_n} - \frac{1}{2} x^\top (\Sigma_n^{-1} - \widetilde{\Sigma}_n^{-1}) x. \quad (2.19)$$

Recall that if $Z \sim \mathcal{N}(0, S)$ then $\mathbb{E}[Z^\top A Z] = \text{tr}(AS)$. Hence

$$\mathbb{E}_{\mathbf{P}_n} \left[\log \frac{\mathbf{P}_n}{\mathbf{Q}_n} \right] = -\frac{1}{2} \left[\log \frac{\det \Sigma_n}{\det \widetilde{\Sigma}_n} + n - \text{tr}(\widetilde{\Sigma}_n^{-1} \Sigma_n) \right].$$

Lemma 4.8 in [18] now yields that

$$\frac{1}{n} \text{tr}(\widetilde{\Sigma}_n^{-1}(\widetilde{A}, \widetilde{A})\Sigma_n(A, A)) \rightarrow \frac{1}{2\pi} \int_{\mathbb{T}} \frac{f}{g} d\lambda du.$$

Hence the asymptotic Kullback-Leibler information of two locally stationary Gaussian processes (as termed by Taniguchi and Kakizawa) with time-varying spectral densities

f and g , respectively, is

$$\begin{aligned}
K_L &= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_{\mathbf{P}_n} \left[\log \frac{d\mathbf{P}_n}{d\mathbf{Q}_n} \right] \\
&= \left(-\frac{1}{2}\right) \lim_{n \rightarrow \infty} \frac{1}{n} \left[\log \frac{\det \Sigma_n}{\det \widetilde{\Sigma}_n} + n - \text{tr} \left(\widetilde{\Sigma}_n^{-1} \Sigma_n \right) \right] \\
&= \left(-\frac{1}{2}\right) \frac{1}{2\pi} \int_{\mathbb{T}} \left(\log \frac{f}{g} + 1 - \frac{f}{g} \right) d\lambda du \\
&= \frac{1}{4\pi} \int_{\mathbb{T}} \left(\log \frac{g}{f} + \frac{f-g}{g} \right) d\lambda du.
\end{aligned}$$

CHAPTER 3

BAYESIAN ESTIMATORS FOR LOCALLY STATIONARY GAUSSIAN PROCESSES

In this chapter we focus on the problem of estimating the trend function, μ_ϑ , and the time-varying spectral density f_ϑ , $\vartheta \in \Theta \subset \mathbb{R}^d$, of a locally stationary Gaussian process under mild regularity conditions. In particular, we consider a wide class of Bayesian estimators and provide consistency, asymptotic normality as well as asymptotic efficiency of this class of estimators. As a by-product, similar results are also obtained for the maximum likelihood estimator; the estimators considered in this chapter are shown to be asymptotically equivalent, i.e., they have the same limit distributions.

3.1 Introduction

The main characteristics of a locally stationary process, e.g., trend and covariance structure, vary smoothly as a function of time and frequency. Then, when restricted to a Gaussian framework, as in our case, a quantitative knowledge about trend and covariance matrix would contribute to a better understanding of the properties of the asymptotic law of the process.

In this chapter, we study the problem of estimating the trend function μ_ϑ and the time-varying spectral density function f_ϑ of a locally stationary Gaussian process $X = (X_{k,n})_{1 \leq k \leq n}$, where $\vartheta \in \Theta \subset \mathbb{R}^d$ and Θ is an open set. The trend function represents the mean while the time-varying spectral density function asymptotically defines the covariance matrix of X . We focus on a class of Bayesian estimators of the parameter ϑ and establish consistency, asymptotic normality as well as asymptotic efficiency of this class of estimators.

As explained in the following section, a locally stationary Gaussian process X is associated with the statistical experiment $\{\mathbf{P}_{\vartheta,n} = \mathcal{N}_n(\mu_\vartheta, \Sigma_{\vartheta,n}) \mid \vartheta \in \Theta\}$. The matrix $\Sigma_{\vartheta,n}$ is asymptotically defined through the function f_ϑ . The assumption of Gaussianity allows us to obtain a convenient expression for the normalized likelihood ratio process $d\mathbf{P}_{\vartheta_n+u_n/\sqrt{n},n}/d\mathbf{P}_{\vartheta_n,n}$, when $u_n \rightarrow u \in \mathbb{R}^d$, $\vartheta_n \in K \subset \Theta$ and K is a compact set. The regular asymptotic behavior of the normalized likelihood ratio process allows us to apply a general method of asymptotic estimation due to Ibragimov and Khasminskii [32] and obtain consistency, asymptotic normality as well as asymptotic efficiency of the Bayesian estimators. More precisely, let $\tilde{\vartheta}_n$ be a Bayesian estimator corresponding to the prior density $d\mathbf{Q}(\cdot)$ and loss function $\ell_n(\cdot)$, where $\ell \in \widetilde{\mathcal{W}}$, then we show that $\tilde{\vartheta}_n$ is consistent, asymptotically normal and efficient, in a non-Bayesian sense, in K for any loss function $w \in \mathcal{W}_p$, i.e., loss functions which possess a polynomial majorant. Both classes of loss functions $\widetilde{\mathcal{W}}$ and \mathcal{W}_p will be properly introduced in the next section.

It is worth mentioning that the same method allows us to obtain similar results for the maximum likelihood estimator of ϑ ; in particular, we show that the Bayesian estimators and the maximum likelihood estimator are asymptotically equivalent, i.e., they have the same limit distributions. All of our results are uniform in ϑ , i.e., they hold over compact sets of the parameter space $\Theta \subset \mathbb{R}^d$.

Unlike Bayesian estimators, the properties of the maximum likelihood estimate of a locally stationary Gaussian process have been studied previously. For instance, Dahlhaus [19] has established the local asymptotic normality (LAN) property of the statistical experiment associated with $X \sim \mathbf{P}_{\vartheta,n} := \mathcal{N}_n(\mu_\vartheta, \Sigma_{\vartheta,n})$. From this result, the asymptotic efficiency of the maximum likelihood estimate $\widehat{\vartheta}_n$ of ϑ_t (true parameter) can be deduced. Indeed, combining Theorem 4.2 of [19] and Definition 2.4 of Millar [39],

we deduce that for any sequence of estimators ρ_n and any bounded function ℓ on \mathbb{R}^d ,

$$\lim_{C \uparrow \infty} \lim_{n \rightarrow \infty} \inf_{T_n} \sup_{\vartheta \in A_n(C, \vartheta_t)} \int \ell(\sqrt{n}(\rho_n - \vartheta)) d\mathbf{P}_{\vartheta, n} = \lim_{C \uparrow \infty} \lim_{n \rightarrow \infty} \sup_{\vartheta \in A_n(C, \vartheta_t)} \int \ell(\sqrt{n}(\widehat{\vartheta}_n - \vartheta)) d\mathbf{P}_{\vartheta, n},$$

where $A_n(C, \vartheta_t) = \{ \vartheta \in \Theta \mid \|\vartheta - \vartheta_t\|_2 \leq Cn^{-1/2} \}$ and the infimum is computed over all the estimators ρ_n available in the experiment $\{\mathbf{P}_{\vartheta, n} \mid \vartheta \in \Theta\}$. As seen in the following section, this result is also valid for loss function $w \in \mathcal{W}_p$.

In this chapter we show that the LAN property holds uniformly over compact subsets of $\Theta \subset \mathbb{R}^d$, i.e., over compact sets of \mathbb{R}^d locally stationary Gaussian processes are *uniformly locally asymptotically normal* (ULAN); the latter is a concept due to Ibragimov and Khasminskii; cf. Definition 2.2, Ch. 2 of [32].

Our contribution in the area of asymptotic estimation consists of applying standard asymptotic theory for regular experiments in a framework in which the assumption of independence is no longer present. In order to overcome the difficulties imposed by the lack of independence between the observations, we base parts of our technical arguments on the work of Dahlhaus [19]. Having used the methods of Ibragimov and Khasminskii, which were not used in [19], we provide new asymptotic results on the problem of estimation of locally stationary Gaussian processes. Particularly, we show the consistency, asymptotic normality and efficiency of a class of Bayesian estimators of the trend and the time-varying spectral density functions.

Recently, some methods developed in [32] have been applied to establish asymptotic properties of adaptive maximum likelihood estimate for ergodic diffusion based models driven by discrete observations, cf. Uchida and Yoshida [47]. In the same spirit, Dachian [15, 16, 17] has conducted a comprehensive study of diverse estimation problems associated with non regular statistical experiments driven by Poisson processes.

In these non regular estimation problems, cf. Ch. 6 of [32], of change-point type,

limit experiments are as a rule non-Gaussian, and efficient estimators are Bayesian rather than maximum likelihood. We believe that the present study of Bayesian estimators in locally stationary Gaussian processes, in regular (Gaussian) limit experiment situations, prepares the way methodologically for treating the more involved case of change-point type problems in locally stationary Gaussian processes.

This chapter is structured as follows. In order to utilize the methods developed in Ch. 3 of Ibragimov and Khasminskii [32] we need to present a limit expression for the properly normalized likelihood ratio process of the model under investigation. We provide such a limit result for locally stationary Gaussian processes in Proposition 3.2 of Section 3.2. Additionally, we provide a series of results, cf. Lemmas 3.11, 3.12, 3.13 and 3.14 in Section 3.3, to establish the regularity of the likelihood ratio process. These results allow us to obtain the claimed asymptotic results for a class of Bayesian estimators, cf. Theorem 3.5 and Theorem 3.6 of subsection 3.2.1, and the maximum likelihood estimator, cf. Theorem 3.8 of subsection 3.2.2. The technicalities required in our proofs are deferred to an Appendix.

3.2 Main results

As explained in the introductory chapter of this work, having proper estimates of the time-varying spectral density function would contribute to having a suitable description of the properties of the entire probability law of a locally stationary Gaussian process.

In this chapter we assume that we observe a realization $X := (X_{k,n})_{1 \leq k \leq n}$ from a locally stationary Gaussian process with true trend function μ and transfer function A° , and we fit a class of locally stationary Gaussian processes with trend function μ_θ , transfer function A_θ° , approximating function A_θ , time-varying spectral density function f_θ and

covariance matrix Σ_ϑ , $\vartheta \in \Theta \subset \mathbb{R}^d$, where Θ is an open set. Notice that for convenience, we have removed the index n from the notation of the covariance function $\Sigma_{\vartheta,n}$.

Throughout this chapter we utilize the following:

Assumption 3.1.

1. There exists a constant κ such that

$$\sup_{k,\lambda} \left| \widetilde{\nabla} \left\{ A_{\vartheta,k,n}^\circ(\lambda) - A_\vartheta \left(\frac{k}{n}, \lambda \right) \right\} \right| \leq \frac{\kappa}{n}$$

where $\widetilde{\nabla} = \nabla_i$ or $\widetilde{\nabla} = \nabla_{ij}$, i.e., $\widetilde{\nabla}$ takes values on the gradient ∇ or Hessian ∇^2 , the derivatives are taken with respect to ϑ .

2. The *true parameter* ϑ_0 is an interior point of Θ .
3. There exists an $\varepsilon > 0$ such that $\varepsilon \geq |A_\vartheta(u, \lambda)| \geq \varepsilon^{-1}$ for all ϑ .
4. $A_\vartheta(u, \lambda)$, $\nabla_i A_\vartheta(u, \lambda)$ and $\nabla_{ij}^2 A_\vartheta(u, \lambda)$ are differentiable in ϑ and u and λ for all i and j ; the mixed derivative $\partial^2/(\partial u \partial \lambda)$ is uniformly continuous.
5. $\mu_\vartheta(u)$, $\nabla_i \mu_\vartheta(u)$ and $\nabla_{ij}^2 \mu_\vartheta(u)$ are differentiable in ϑ and u for all i and j with uniformly continuous derivatives.
6. $A_\vartheta(u, \lambda)$ is twice differentiable in u with uniformly bounded derivative for all ϑ .
7. The trend function A_ϑ° also satisfies the assumptions above.

Proposition 3.2 below establishes that for any compact set $K \subset \Theta$, the likelihood ratio process $d\mathbf{P}_{\vartheta_n+u_n/\sqrt{n},n}/d\mathbf{P}_{\vartheta_n,n}$ converges weakly to a normal distribution with parameters $(-\frac{1}{2}u^\top \Gamma_{\vartheta_0} u, u^\top \Gamma_{\vartheta_0} u)$, provided $\vartheta_n \rightarrow \vartheta_0 \in K$ and $u_n \rightarrow u \in \mathbb{R}^d$. This is the first among a series of results, the rest can be found in Section 3.3, about the regularity of the statistical experiment defined by observations from a locally stationary Gaussian process. These results allow us to apply the general method of Ibragimov and Khasminskii

and obtain the main results of this paper. Before stating Proposition 3.2, we need to introduce some notation.

For instance, from now on, \mathcal{K}_Θ denotes the class of compact sets of Θ ; for any $d \times d$ matrix A and $x, y \in \mathbb{R}^d$, $\langle x, y \rangle_A$ is short notation for $x^\top A y$; $Y_n \xrightarrow{\mathbf{P}} 0$ means that Y_n converges to 0 in \mathbf{P} -probability while $\mathfrak{L}(Y_n | \mathbf{P}) \Rightarrow Y$ means that the law of Y_n , under the probability \mathbf{P} , converges weakly to Y ; $\|x\|_2$ denotes the Euclidean norm of $x \in \mathbb{R}^d$.

The log-likelihood function of the parameter $\vartheta \in \Theta$ is given by

$$\mathcal{L}_n(\vartheta; X) = \frac{1}{2} \log(2\pi) + \frac{1}{2n} \log \det \Sigma_\vartheta + \frac{1}{2n} (X - \mu_\vartheta)^\top \Sigma_\vartheta^{-1} (X - \mu_\vartheta). \quad (3.1)$$

For convenience we sometimes write $\mathcal{L}_n(\vartheta)$.

For $\vartheta \in \Theta$ and $\Sigma_\vartheta = \Sigma_n(A_\vartheta, A_\vartheta)$, cf. (1.5), we set

$$C_\vartheta^{(i)} = \nabla_i \Sigma_\vartheta = \Sigma_n(\nabla_i A_\vartheta, A_\vartheta) + \Sigma_n(A_\vartheta, \nabla_i A_\vartheta), \quad i = 1, \dots, d, \quad (3.2)$$

$$D_\vartheta^{(i,j)} = \frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \Sigma_\vartheta, \quad i, j = 1, \dots, d. \quad (3.3)$$

The gradient of the log-likelihood function $\mathcal{L}_n(\vartheta)$ is given by

$$\begin{aligned} \nabla_i \mathcal{L}_n(\vartheta) &= \frac{1}{2n} \text{tr} \left\{ \Sigma_\vartheta^{-1} C_\vartheta^{(i)} \right\} - \frac{1}{2n} (X - \mu_\vartheta)^\top \Sigma_\vartheta^{-1} C_\vartheta^{(i)} \Sigma_\vartheta^{-1} (X - \mu_\vartheta) \\ &\quad - \frac{1}{n} (\nabla_i \mu_\vartheta)^\top \Sigma_\vartheta^{-1} (X - \mu_\vartheta), \quad i = 1, \dots, d. \end{aligned} \quad (3.4)$$

For any $\vartheta \in \Theta$ we define the $d \times d$ positive-definite matrix

$$\Gamma_\vartheta = \frac{1}{4\pi} \int_{\mathbb{T}} \left(\frac{\partial}{\partial \vartheta} \log f_\vartheta(u, \lambda) \right) \left(\frac{\partial}{\partial \vartheta} \log f_\vartheta(u, \lambda) \right)^\top d\lambda du. \quad (3.5)$$

Proposition 3.2. *Let $K \in \mathcal{K}_\Theta$ and suppose that Assumption 3.1 holds. For $\varphi_n(\vartheta_n) = n^{-1/2} \Gamma_{\vartheta_n}^{-1/2}$ and arbitrary sequences $\vartheta_n \in K$, $u_n \rightarrow u \in \mathbb{R}^d$ such that $\vartheta_n + \varphi_n(\vartheta_n)u_n \in K$, the following result is valid:*

$$\log \frac{d\mathbf{P}_{\vartheta_n + \varphi_n(\vartheta_n)u_n, n}}{d\mathbf{P}_{\vartheta_n, n}}(X) - \langle S_n(\vartheta), u \rangle + \frac{1}{2} \|u\|_2^2 \xrightarrow{\mathbf{P}_{\vartheta_n, n}} 0, \quad (3.6)$$

where $S_n(\vartheta_n) = -\sqrt{n}\Gamma_{\vartheta_n}^{-1/2}\nabla\mathcal{L}_n(\vartheta_n)$, and

$$\mathfrak{L}(S_n(\vartheta_n) \mid \mathbf{P}_{\vartheta_n,n}) \Rightarrow \mathcal{N}(0, I_{d \times d}). \quad (3.7)$$

Proof. Since K is compact it suffices to show the validity of eqs. (3.6) and (3.7) for convergent sequences. Let $\vartheta_n \rightarrow \vartheta_0 \in K$. It is also immediate that $\vartheta_n + \varphi(\vartheta_n)u_n \in K$ for n large enough.

From eq. (3.1) follows that for all $\vartheta \in \Theta$, $\log d\mathbf{P}_{\vartheta,n}(X) = -n\mathcal{L}_n(\vartheta)$. Using this relation and the previous paragraph we have an expression for the log-likelihood ratio of $\mathbf{P}_{\vartheta_n+\varphi_n(\vartheta_n)u_n,n}$ w.r.t. $\mathbf{P}_{\vartheta_n,n}$. Namely,

$$\log \frac{d\mathbf{P}_{\vartheta_n+\varphi_n(\vartheta_n)u_n,n}}{d\mathbf{P}_{\vartheta_n,n}}(X) = -n [\mathcal{L}_n(\vartheta_n + \varphi_n(\vartheta_n)u_n) - \mathcal{L}_n(\vartheta_n)].$$

From a second-order Taylor expansion and the mean value theorem we obtain that

$$\mathcal{L}_n(\vartheta_n + \varphi_n(\vartheta_n)u_n) - \mathcal{L}_n(\vartheta_n) = \frac{1}{\sqrt{n}} u_n^\top \nabla \mathcal{L}_n(\vartheta_n) + \frac{1}{2n} u_n^\top \nabla^2 \mathcal{L}(\vartheta_n^*) u_n.$$

where $\|\vartheta_n^* - \vartheta_0\|_2 \leq n^{-1/2}$. Then,

$$\log \frac{d\mathbf{P}_{\vartheta_n+\varphi_n(\vartheta_n)u_n,n}}{d\mathbf{P}_{\vartheta_n,n}}(X) = \left\langle u, -\sqrt{n}\Gamma_{\vartheta_0}^{-1}\nabla\mathcal{L}_n(\vartheta_n) \right\rangle_{\Gamma_{\vartheta_0}} - \frac{1}{2}\|u_n\|_{\Gamma_{\vartheta_0}}^2 + \Psi_n(\vartheta_n),$$

where

$$\Psi_n(\vartheta_n) = \left\langle u_n - u, -\sqrt{n}\Gamma_{\vartheta_0}^{-1}\nabla\mathcal{L}_n(\vartheta_n) \right\rangle_{\Gamma_{\vartheta_0}} - \frac{1}{2}\|u_n\|_{A_n(\vartheta_n^*)}^2,$$

$$A_n(\vartheta_n^*) = \nabla^2 \mathcal{L}_n(\vartheta_n^*) - \Gamma_{\vartheta_0}.$$

In order to prove that (3.7) holds, it suffices to show that for any $w \in \mathbb{R}^d$, $\langle S_n(\vartheta_n), w \rangle_{\Gamma_{\vartheta_0}}$ converges weakly to $\langle Z, w \rangle_{\Gamma_{\vartheta_0}}$, cf. Theorem 6 of Ibragimov and Khasminskii [32]. From Lemma 3.18 below we know that $\mathfrak{L}(\langle u_n, S_n(\vartheta_n) \rangle_{\Gamma_{\vartheta_0}} \mid \mathbf{P}_{\vartheta_n}) \Rightarrow \mathfrak{L}(\langle u, Z \rangle_{\Gamma_{\vartheta_0}})$ provided $u_n \rightarrow u$.

Next, set $w_n = u_n - u$ and observe that Lemma 3.18 now yields that $\langle w_n, S_n(\vartheta_n) \rangle_{\Gamma_{\vartheta_0}}$ tends to 0 in $\mathbf{P}_{\vartheta_n, n}$ -probability. Combining Lemmas 3.16 and 3.17 we obtain that $A_n(\vartheta_n^*) \rightarrow 0$ in $\mathbf{P}_{\vartheta_n, n}$ -probability. Consequently, $\Psi_n(\vartheta_n) \xrightarrow{\mathbf{P}_{\vartheta_n, n}} 0$. Since $\|u_n\|_2^2 \rightarrow \|u\|_2^2$, then $\|u_n\|_{\Gamma_{\vartheta_0}}^2 \rightarrow \|u\|_{\Gamma_{\vartheta_0}}^2$. This completes the proof. \square

Any statistical experiment fulfilling eqs. (3.6) and (3.7) is called *uniformly locally asymptotically normal* (ULAN) in accordance with Definition 2.2 of Ibragimov and Khasminskii [32]. In Ch. 1 of [32], it has been established that for experiments given by regular i.i.d. observations with nondegenerate asymptotic Fisher information matrix $I(\vartheta)$ in $\Theta \subset \mathbb{R}$, the corresponding family of distributions is uniformly asymptotically normal in any compact set $K \subset \Theta$. For nonstationary processes the validity of the LAN property has been proved in some cases; locally stationary Gaussian processes are LAN according to Theorem 2.1 of Dahlhaus [19]; Hirukawa and Taniguchi have proved the LAN property of a family of non-Gaussian locally stationary processes. Proposition 3.2 is a generalization of Dahlhaus' result which allows us to show the consistency, asymptotic normality and efficiency of a class of Bayes-type estimators and the maximum likelihood as explained in the following subsections.

3.2.1 Bayesian estimators

In this section we establish the consistency, asymptotic normality and efficiency of $\widetilde{\vartheta}_n$, a Bayesian estimator corresponding to the prior density $d\mathbf{Q}(\cdot)$ and loss function $\ell(n^{1/2}x)$. To this end, we now introduce some definition and notation.

Recall that the Bayesian estimator $\widetilde{\vartheta}_n$ with respect to the prior density $d\mathbf{Q}(\cdot)$ and loss

function $\ell(\cdot)$ is the solution of

$$\inf_{y \in \Theta} \int \ell(u - y) d \Pr(u) = \int \ell(u - \tilde{\vartheta}_n) d \Pr(u),$$

where \Pr denotes the posterior distribution of ϑ . In this work we consider the class of loss functions \mathcal{W}_p which is defined as follows.

Definition 3.3. The class of loss functions W is defined on the set $\Theta \times \Theta$ and satisfies the following properties:

1. $W(u, v) = w(u - v)$.
2. $w(u)$ is defined and nonnegative on \mathbb{R}^p ; $w(0) = 0$, w is continuous at $u = 0$ (but is not identically zero).
3. $w(u) = w(-u)$.
4. (a) The sets $\{u : w(u) < c\}$ are convex sets $\forall c > 0$.
 (b) The sets $\{u : w(u) < c\}$ are convex sets $\forall c > 0$ and are bounded $\forall c > 0$ sufficiently small.

\mathcal{W} is the class of functions satisfying 1-4(a); \mathcal{W}' is the class of functions satisfying 1-4(b). The notation \mathcal{W}_p (\mathcal{W}'_p) will be used for the set of functions $w \in \mathcal{W}$ (\mathcal{W}') which possess a polynomial majorant.

In this section we consider loss functions of the form

$$\ell_n(x) = \ell(n^{1/2} x).$$

Additionally, it will be assumed that $\ell(x) \in \mathcal{W}'_p$ and, moreover, there exists a $\gamma > 0$ such that for all $a \geq a_0$,

$$\inf_{\|s\| \geq a} \ell(s) \geq \sup_{\|s\| \leq a^\gamma} \ell(s).$$

A class of such functions will be denoted by $\widetilde{\mathcal{W}}$. The prior density $d\mathbf{Q}(\cdot)$ will be assumed to be a function continuous and positive in Θ whose growth is at most polynomial and which is not necessarily integrable.

The following definition introduces a class of matrices whose role in the following consists of providing a suitable class of normalizing matrices for the likelihood ratio process, $\mathbf{P}_{\vartheta+h}/\mathbf{P}_{\vartheta}$. For instance, the asymptotic efficiency of the maximum likelihood estimate $\widehat{\vartheta}_n$ depends on the *regularity* of the normalizing matrix $\varphi_n(\vartheta)$, cf. Corollary 3.10 below.

Definition 3.4 (Class $\Phi(K)$). We write that the family $\varphi_n(\vartheta) \in \Phi(K)$ if $\varphi_n(\vartheta)$ is such that for $\vartheta_i \in \Theta$ there exists the limit

$$\lim_{n \rightarrow \infty} \varphi_n^{-1}(\vartheta_2) \varphi_n(\vartheta_1) = B(\vartheta_1, \vartheta_2),$$

where this convergence is uniform in $\vartheta_i \in K \subset \Theta$ and the matrix $B(\vartheta_1, \vartheta_2)$ is continuous in ϑ_1 .

Recall that in Proposition 3.2 we utilized $\varphi_n(\vartheta) = n^{-1/2} \Gamma_{\vartheta}^{-1/2}$ as the normalizing matrix to establish the ULAN property of locally stationary Gaussian processes. It is easily seen that in this case, $\varphi_n(\vartheta) \in \Phi(K)$ uniformly in ϑ .

Applying Theorem 2.1, Ch. 3, p. 179, of [32], we have the following:

Theorem 3.5. *Let $\widetilde{\vartheta}_n$ be a Bayesian estimator corresponding to the prior density $d\mathbf{Q}(\cdot)$ and loss function $\ell(n^{1/2}\cdot)$, $\ell \in \widetilde{\mathcal{W}}$. Suppose that Assumption 3.1 holds. Then, uniformly in ϑ ,*

$$\begin{aligned} \widetilde{\vartheta}_n &\xrightarrow{\mathbf{P}_{\vartheta,n}} \vartheta, \\ \sqrt{n}(\vartheta - \widetilde{\vartheta}_n) &\Rightarrow \mathcal{N}(0, \Gamma_{\vartheta}^{-1}) \\ \lim_{C \uparrow \infty} \lim_{n \uparrow \infty} \inf_{\rho_n} \sup_{\nu \in A_n(C, \vartheta)} \int \omega(\sqrt{n}(\rho_n - \nu)) d\mathbf{P}_{\nu,n} &= \lim_{C \uparrow \infty} \lim_{n \uparrow \infty} \sup_{\nu \in A_n(C, \vartheta)} \int \omega(\sqrt{n}(\widetilde{\vartheta}_n - \nu)) d\mathbf{P}_{\nu,n}, \end{aligned}$$

where $A_n(C, \vartheta) = \{ \nu \in \Theta \mid \|\vartheta - \nu\|_2 \leq Cn^{-1/2} \}$, the infimum is taken over all the estimators ρ_n and $\omega \in \mathcal{W}_p$.

Proof. In view of Lemmas 3.11, 3.12, 3.13 and 3.14, which can be found in the appendix, the conditions of Theorem 3.2.1 of Ibragimov and Khasminskii [32] are satisfied. \square

The following result shows that under the conditions imposed in this section, the Bayesian estimators corresponding to different prior densities and loss functions are asymptotically equivalent among themselves and are also equivalent (cf. Theorem 3.8) to the maximum likelihood estimator of ϑ .

Theorem 3.6. For any $K \in \mathcal{K}_\Theta$ and under the assumptions of Theorem 3.5, for any $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \sup_{\vartheta \in K} \mathbf{P}_\vartheta \left\{ \|\sqrt{n}(\widetilde{\vartheta}_n - \vartheta) - S_n(\vartheta)\|_2 > \varepsilon \right\} = 0,$$

where $S_n(\vartheta) = -\sqrt{n}\Gamma_\vartheta^{-1/2} \nabla \mathcal{L}_n(\vartheta)$.

Proof. In view of Lemmas 3.11, 3.12, 3.13 and 3.14, which can be found in the appendix, this result is a consequence of Theorem 3.2.2 of Ibragimov and Khasminskii [32]. \square

In particular, this proposition states that for any convergent sequence $\{\vartheta_k\} \subset K$, where K is a closed set whose diameter is finite, e.g., $K(\vartheta_t) = \{ \vartheta \in \Theta \mid \|\vartheta - \vartheta_t\|_2 \leq Cn^{-1/2} \}$, $\sqrt{n}(\widetilde{\vartheta}_n - \vartheta_k) - S_n(\vartheta_k)$ converges to zero in $\mathbf{P}_{\vartheta_k, n}$ -probability.

We have shown in Lemma 3.18 that under $\mathbf{P}_{\vartheta_n, n}$, $S_n(\vartheta_n)$ weakly converges to $Z = \mathcal{N}(0, \Gamma_{\vartheta_0}^{-1})$, provided $\vartheta_n \rightarrow \vartheta_0$. An application of Slutsky's theorem now yields that for any sequence $\{\vartheta_k\} \subset K$, under \mathbf{P}_{ϑ_k} , $\sqrt{n}(\widetilde{\vartheta}_n - \vartheta_k)$ also converges weakly to Z .

Moreover, by an application of the second continuous mapping theorem, cf. Theorem 5.5 in Billingsley [6], we get that for any sequence $a_n \rightarrow \tau$

$$\mathbf{P}_\vartheta\{ \sqrt{n}\|\widetilde{\vartheta}_n - \vartheta\|_2 \geq a_n \} = o(1) + \mathbf{P}_\vartheta\{ Z^\top Z \geq \tau \}.$$

Chebyshev's inequality now yields

$$\mathbf{P}_\vartheta\{ Z^\top Z \geq \tau \} \leq \frac{\mathbb{V}[Z^\top Z]}{\tau^2} = \frac{\text{tr}(\Gamma_\vartheta^{-2})}{\tau^2}.$$

Since for any $d \times d$ matrix M , $|\text{tr}(M^2)| \leq \|M\|^2$, and $\|M^{-1}\| = 1/\sigma_d$, where σ_d is the smallest eigenvalue of M , and $\|M\| = [\text{tr}(M M^*)]^{1/2}$ is the Frobenius norm of M , it follows that

$$\left| \frac{\text{tr}(\Gamma_\vartheta^{-2})}{\tau^2} \right| \leq \left(\frac{1}{\sigma_d(\vartheta) \tau} \right)^2,$$

where $\sigma_d(\vartheta)$ is the smallest eigenvalue of Γ_ϑ . Because for any $\vartheta \in \Theta$, $\sigma_d(\vartheta)$ is bounded from below, if we choose τ large enough we have the following:

Corollary 3.7. *Suppose that Assumption 3.1 holds. Then, for any $K \in \mathcal{K}_\Theta$ there exists a number τ large enough such that for any $\varepsilon > 0$ the following inequality*

$$\sup_{\vartheta \in K} \mathbf{P}_\vartheta \left\{ \sqrt{n}\|\widetilde{\vartheta}_n - \vartheta\|_2 > \tau \right\} \leq \varepsilon,$$

is valid.

This corollary proves the existence of \sqrt{n} -estimators of the time-varying spectral density function f_ϑ which are uniformly bounded in probability. This result is crucial to treat the estimation of f_ϑ from the view point of LeCam theory of statistical experiments; this problem will be addressed in the next chapter.

3.2.2 Maximum likelihood estimator

A maximum likelihood estimate $\widehat{\vartheta}_n$ of the parameter ϑ based on observations X will be understood as one of the values y which maximizes with respect to $y \in \Theta$ the likelihood function $\partial \mathbf{P}_y^n(X)/\partial \nu_n$, where ν_n is the measure with respect to which all the measures \mathbf{P}_y^n , $y \in \Theta$, are absolutely continuous. Disregarding events of \mathbf{P}_{ϑ}^n -probability 0, $\widehat{\vartheta}_n$ maximizes $\partial \mathbf{P}_y^n(X)/\partial \mathbf{P}_{\vartheta_0}^n$, where ϑ_0 is the true value of the parameter.

In this section we consider $\widehat{\vartheta}_n$ as the maximum likelihood estimate of ϑ . The following theorem presents a method for proving the asymptotic normality of $\widehat{\vartheta}_n$. This theorem is a direct consequence of Theorem 3.1.2 of Ibragimov and Khasminskii [32]. The latter theorem is of interest on its own; we have chosen to utilize Theorem 3.1.2 since its conditions are in line with the mild assumptions that we have imposed on the trend and the approximating functions, μ_{ϑ} and A_{ϑ} , respectively, cf. Definition 1.1. We should mention that under more restrictive assumptions Theorem 1.8.1 of [32] has shown useful in providing stronger results than Theorem 3.1.2 in i.i.d. settings.

Theorem 3.8. *Suppose that Assumption 3.1 holds. Then, for any set $K \subset \mathcal{K}_{\Theta}$ and any $\varepsilon > 0$,*

$$\lim_{n \rightarrow \infty} \sup_{\vartheta \in K} \mathbf{P}_{\vartheta}\{\|\sqrt{n}(\widehat{\vartheta}_n - \vartheta) - S_n(\vartheta)\|_2 > \varepsilon\} = 0, \quad (3.8)$$

where $S_n(\vartheta) = -\sqrt{n}\Gamma_{\vartheta_0}^{-1/2} \nabla \mathcal{L}_n(\vartheta)$.

Proof. In view of Lemmas 3.11, 3.12, 3.13 and 3.14, which can all be found in the appendix, this result is a consequence of Theorem 3.1.2 of Ibragimov and Khasminskii [32]. □

Since asymptotically $\widehat{\vartheta}_n$ (maximum likelihood estimate) and $\widetilde{\vartheta}_n$ (Bayesian estimators) are equivalent, the following result is expected and can be proved following the same arguments used to show Corollary 3.7.

Corollary 3.9. *Suppose that Assumption 3.1 holds. Then, for any $K \in \mathcal{K}_\Theta$ there exists a number τ large enough such that for any $\varepsilon > 0$ the following inequality*

$$\sup_{\vartheta \in K} \mathbf{P}_\vartheta \left\{ \sqrt{n} \|\widehat{\vartheta}_n - \vartheta\|_2 > \tau \right\} \leq \varepsilon,$$

is valid.

Applying Corollary 1.1, Ch. 3, p. 177, of Ibragimov and Khasminskii [32], we have the following:

Corollary 3.10. *For any $K \in \mathcal{K}_\Theta$ and under the assumptions of Theorem 3.8, the maximum likelihood estimate $\widehat{\vartheta}_n$ is asymptotically efficient in K for the loss function $w(\varphi_n^{-1}(\vartheta_0)x)$ for any $\vartheta_0 \in K$ and $w \in \mathcal{W}_p$.*

3.3 Appendix

In this section, $\mathbf{E}_\vartheta[X]$ and $\mathbf{V}_\vartheta[X]$ denote the expectation and variance of X w.r.t. the law \mathbf{P}_ϑ ; \mathcal{K}_Θ denotes the class of compact sets of Θ ; $\|x\|_2$ denotes the Euclidean norm of $x \in \mathbb{R}^d$; for any $n \times n$ matrix M , $\|M\| = (\max. \text{characteristic root of } M^* M)^{1/2}$, where M^* denotes the conjugate transpose of M .

Lemma 3.11 (Condition N1 in [32]). *For some nondegenerate matrix $\varphi_n(\vartheta)$, for any set $K \subset \mathcal{K}_\Theta$ and arbitrary sequences $\vartheta_n \in K$, $\vartheta_n + \varphi_n(\vartheta_n)u_n \in K$, $u_n \rightarrow u$, $n \rightarrow \infty$, the representation*

$$\begin{aligned} \mathcal{Z}_{n,\vartheta_n}(u_n) &= \frac{d\mathbf{P}_{\vartheta_n + \varphi_n(\vartheta_n)u_n,n}}{d\mathbf{P}_{\vartheta_n,n}}(X) \\ &= \exp \left\{ \langle S_n(\vartheta_n), u \rangle - \frac{1}{2} \|u\|^2 + \psi_n(u_n, \vartheta_n) \right\} \end{aligned}$$

is valid; here $\mathcal{L}(S_n(\vartheta_n) \mid \mathbf{P}_{\vartheta_n,n}) \Rightarrow \mathcal{N}(0, I_{p \times p})$ as $n \rightarrow \infty$ and the sequence $\psi_n(u_n, \vartheta_n)$ converges to zero in $\mathbf{P}_{\vartheta_n,n}$ -probability.

Proof. This is precisely the content of Proposition 3.2. Thus, let $\varphi_n(\vartheta_n) = n^{-1/2} \Gamma_{\vartheta_n}$ and apply the aforementioned proposition. \square

Lemma 3.12 (Condition N2 in [32]). *For any set $K \subset \mathcal{K}_\Theta$,*

$$\lim_{n \rightarrow \infty} \sup_{\vartheta \in K} \text{tr} \{ \varphi_n(\vartheta) \varphi_n^\top(\vartheta) \} = 0.$$

Proof. Since for any $\vartheta \in \Theta$,

$$\text{tr}(\Gamma_\vartheta) = \frac{1}{4\pi} \sum_{i=1}^d \int_{\mathbb{T}} \left(\frac{\partial}{\partial \vartheta_i} \log f_\vartheta \right)^2 d\lambda du = \frac{1}{4\pi} \sum_{i=1}^d \int_{\mathbb{T}} \left(\frac{\partial / (\partial \vartheta_i) \vartheta_i f_\vartheta}{f_\vartheta} \right)^2 d\lambda du,$$

we use the boundedness away from zero of f_ϑ as well as the boundedness of ∇A_ϑ to get that for some $\varepsilon > 0$ and $\kappa < \infty$

$$|\text{tr}(\Gamma_\vartheta)| \leq \frac{\varepsilon}{4\pi} \sum_{i=1}^d \int_{\mathbb{T}} \left(\frac{\partial}{\partial \vartheta_i} f_\vartheta \right)^2 d\lambda du \leq \frac{d \kappa^2}{4\pi \varepsilon}.$$

Observe that for any $\vartheta \in \Theta$,

$$|\text{tr} \{ \varphi_n(\vartheta) \varphi_n^\top(\vartheta) \}| \leq n^{-1} |\text{tr}(\Gamma_\vartheta)|^2 \leq n^{-1} \left(\frac{d \kappa^2}{4\pi \varepsilon} \right)^2.$$

Let $n \rightarrow \infty$ to conclude the proof. \square

Lemma 3.13 (Condition N3 in [32]). *Let $U_n(\vartheta) = \varphi_n^{-1}(\vartheta)(\Theta - \vartheta)$. For any set $K \in \mathcal{K}_\Theta$ and some $\beta > 0$ and $m > 0$, $B = B(K)$, $a = a(K)$,*

$$\sup_{\vartheta \in K} \sup_{\substack{v, w \in U_n(\vartheta) \\ |v| < R, |w| < R}} \|w - v\|^{-\beta} \mathbb{E}_\vartheta \left| \mathfrak{Z}_{n,\vartheta}^{1/m}(w) - \mathfrak{Z}_{n,\vartheta}^{1/m}(v) \right|^m \leq B(1 + R^a).$$

Proof. We show that this condition holds with $\beta = m = 2$. This result is a consequence of Lemma 1.1, Ch. 3, p. 178, of Ibragimov and Khasminskii [32]. In order to utilize this lemma, we begin by showing that the statistical experiment $\mathbb{F}_n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{ \mathbf{P}_{\vartheta,n}, \vartheta \in \Theta \})$ is regular. Here $\mathbf{P}_{\vartheta,n} = \mathcal{N}_n(\mu_\vartheta, \Sigma_\vartheta)$. Then, we show the boundedness of

$$\sup_{\vartheta \in K} \sup_{\substack{\|u\| < R, \\ \vartheta + u \in \Theta}} \left\| I_n^{-1/2}(\vartheta) I_n(\vartheta + u) I_n^{-1/2}(\vartheta) \right\|, \quad (3.9)$$

where $I_n(\vartheta)$ is the Fisher's information matrix associated with the experiment \mathbb{F}_n .

Note that $d\mathbf{P}_{\vartheta,n}$ is continuous on Θ . Observe also that the function $(d\mathbf{P}_{\vartheta,n})^{1/2}$ is differentiable in $L_2(dx)$ with derivative $\psi_\vartheta(x) = (1/2)[d\mathbf{P}_{\vartheta,n}(x)]^{1/2} \nabla_\vartheta \log d\mathbf{P}_{\vartheta,n}(x)$. Thus the Fisher's information matrix for \mathbb{F}_n is given by

$$I_n(\vartheta) = \mathbb{E}_\vartheta \left[\nabla_\vartheta \log d\mathbf{P}_{\vartheta,n}(x) (\nabla_\vartheta \log d\mathbf{P}_{\vartheta,n}(x))^\top \right] = -\mathbb{E}_\vartheta \left[\nabla_\vartheta^2 \log d\mathbf{P}_{\vartheta,n}(x) \right].$$

The latter follows because normal distributions belong to an exponential family. Thus \mathbb{F}_n is a regular experiment; cf. p. 65 in [32].

Since K is compact it suffices to verify (3.9) for convergent sequences. Let $\vartheta_n \rightarrow \vartheta_0 \in K$ and $u_n \rightarrow u \in \mathbb{R}^d$. As above, $\varphi_n(\vartheta_n) = n^{-1/2} I_{d \times d}$. Let $R = n^{-1/2}$ and observe that the requirement that $\|u_n\|_2 < n^{-1/2}$ implies that $\vartheta_n + u_n \in K$ for n large enough. Since $I_n(\vartheta_n)$ is a Hermitian positive definite matrix, from Proposition V 1.8 and Theorem X 1.1 of Bhatia [5],

$$\begin{aligned} \left\| I_n^{-1/2}(\vartheta_n) I_n(\vartheta_n + u_n) I_n^{-1/2}(\vartheta_n) \right\| &\leq \left\| I_n^{-1}(\vartheta_n) \right\| \|I_n(\vartheta_n + u_n)\| \\ &= \left\| \left(\mathbb{E}_{\vartheta_n} [\nabla^2 \mathcal{L}_n(\vartheta_n)] \right)^{-1} \right\| \left\| \mathbb{E}_{\vartheta_n} [\nabla^2 \mathcal{L}_n(\vartheta_n + u_n)] \right\|. \end{aligned}$$

The latter follows because $\log d\mathbf{P}_{\vartheta,n}(x) = -n\mathcal{L}_n(\vartheta)$ for any $\vartheta \in \Theta$ and therefore, $I_n(\vartheta) = n \mathbb{E}_\vartheta [\nabla^2 \mathcal{L}_n(\vartheta)]$.

From Lemma 3.17 and Lemma 3.16 below we deduce that for any $i, j = 1, \dots, d$,

$$\begin{aligned} \left| \left(\mathbb{E}_{\vartheta_n} [\nabla^2 \mathcal{L}_n(\vartheta_n + u_n)] - \Gamma_{\vartheta_0+u} \right)_{i,j} \right| &\leq \left| \left(\mathbb{E}_{\vartheta_n} [\nabla^2 \mathcal{L}_n(\vartheta_n + u_n)] - \mathbb{E}_{\vartheta_n} [\nabla^2 \mathcal{L}_n(\vartheta_0 + u)] \right)_{i,j} \right| \\ &\quad + \left| \left(\mathbb{E}_{\vartheta_n} [\nabla^2 \mathcal{L}_n(\vartheta_0 + u)] - \Gamma_{\vartheta_0+u} \right)_{i,j} \right| = O(n^{-2/3} \log^4(n)). \end{aligned} \quad (3.10)$$

The rate $n^{-2/3} \log^4(n)$ is taken from Lemma 4.8 of Dahlhaus [19]. This is one of the two parts of the paper in which we need the Assumption 3.1.5; this assumption is also crucial in establishing Lemma 3.18.

Therefore,

$$\begin{aligned} \left\| \mathbb{E}_{\vartheta_n} [\nabla^2 \mathcal{L}_n(\vartheta_n + u_n)] - \Gamma_{\vartheta_0+u} \right\| &\leq \sup_j \sum_i \left| \left(\mathbb{E}_{\vartheta_n} [\nabla^2 \mathcal{L}_n(\vartheta_n + u_n)] \right. \right. \\ &\quad \left. \left. - [\Gamma_{\vartheta_0+u}]_{i,j} \right) \right| = O(d n^{-4/3} \log^8(n)). \end{aligned}$$

Since $\|\Gamma_{\vartheta_0+u}\|$ is uniformly bounded by constants, cf. Lemma A.4 of Dahlhaus [19], the quantity $\left\| \mathbb{E}_{\vartheta_n} [\nabla^2 \mathcal{L}_n(\vartheta_n + u_n)] \right\| = O(d n^{-4/3} \log^8(n)) + \|\Gamma_{\vartheta_0+u}\|$ is uniformly bounded.

Next we show that $\left\| \left(\mathbb{E}_{\vartheta_n} [\nabla^2 \mathcal{L}_n(\vartheta_n)] \right)^{-1} \right\|$ is bounded. First note that

$$\begin{aligned} \left\| \left(\mathbb{E}_{\vartheta_n} [\nabla^2 \mathcal{L}_n(\vartheta_n)] \right)^{-1} \right\| &= \sup_{\|x\|=1} \frac{x^\top x}{x^\top \mathbb{E}_{\vartheta_n} [\nabla^2 \mathcal{L}_n(\vartheta_n)] x} \\ &= \left(\inf_{\|x\|=1} x^\top \mathbb{E}_{\vartheta_n} [\nabla^2 \mathcal{L}_n(\vartheta_n)] x \right)^{-1}. \end{aligned}$$

Since for any $x \in \mathbb{R}^d$ and any $d \times d$ matrix M , $|x^\top M x| \leq \|x\|_2^2 \|M\|$ from eq. (3.10) we deduce that

$$\inf_{\|x\|=1} x^\top \left(\mathbb{E}_{\vartheta_n} [\nabla^2 \mathcal{L}_n(\vartheta_n)] - \Gamma_{\vartheta_0} \right) x \geq -O(n^{-1/3} \log^8(n)).$$

Therefore,

$$\left\| \left(\mathbb{E}_{\vartheta_n} [\nabla^2 \mathcal{L}_n(\vartheta_n)] \right)^{-1} \right\| \leq \left(\inf_{\|x\|=1} x^\top \Gamma_{\vartheta_0} x - O(n^{-1/3} \log^8(n)) \right)^{-1}.$$

Since $\inf_{\|x\|=1} x^\top \Gamma_{\vartheta_0} x$ is bounded from below, our claim follows. This completes the proof. \square

Lemma 3.14 (Condition N4 in [32]). *For any set $K \subset \mathcal{K}_\Theta$ and any $N > 0$, there exists an $n_0(N, K)$ such that*

$$\sup_{\vartheta \in K} \sup_{n_0 < N} \sup_{u \in U_n(\vartheta)} \|u\|_2^N \mathbb{E}_{\vartheta} \mathfrak{Z}_{n,\vartheta}^{1/2}(w) < \infty.$$

Proof. Recall that $U_n(\vartheta) = \sqrt{n}(\Theta - \vartheta)$. It suffices to verify the condition for convergent sequences, $\vartheta_n \rightarrow \vartheta_0 \in K$. Furthermore, since $\sqrt{x} \leq (1+x)/2$ for all $x \geq 0$, it is sufficient

to verify that for n large enough the following condition holds:

$$\sup_{u \in \sqrt{n}(\Theta - \vartheta_n)} \|u\|_2^N \mathbb{E}_{\vartheta_n} \mathfrak{Z}_{n, \vartheta_n}(u) < \infty.$$

To this end observe first that $u \in \sqrt{n}(\Theta - \vartheta_n)$ implies that there is a $\vartheta_u \in \Theta$ such that $u = \sqrt{n}(\vartheta_u - \vartheta_n)$. Thus $\|u\|_2 = \sqrt{n}\|\vartheta_u - \vartheta_n\| \leq \sup\{\sqrt{n}\|\vartheta_1 - \vartheta_2\| : \vartheta_1, \vartheta_2 \in \Theta\} =: \text{diam}(\Theta)$.

Then note that

$$\|u\|_2^N \mathbb{E}_{\vartheta_n} \mathfrak{Z}_{n, \vartheta_n}(u) = \|u\|_2^{N+2} \|u\|^{-2} \mathbb{E}_{\vartheta_n} \mathfrak{Z}_{n, \vartheta_n}(u) \leq (\text{diam}(\Theta))^{N+2} \text{Const.}$$

The latter follows by an application of Lemma 3.13. Since $\text{diam}(\Theta)$ is finite, the proof is completed. \square

Lemma 3.15. *If $\vartheta_n \rightarrow \vartheta_0$ then $\|\Sigma_{\vartheta_n} - \Sigma_{\vartheta_0}\| \rightarrow 0$ as $n \rightarrow \infty$.*

Proof. For any $\vartheta \in \Theta$ we have $\Sigma_n = \left\{ \int_{-\pi}^{\pi} \mathbf{e}^{i\lambda(j-k)} A_{\vartheta, j, n}^{\circ}(\lambda) \bar{A}_{\vartheta, k, n}^{\circ}(\lambda) d\lambda \right\}$, where $j, k = 1, \dots, n$; cf. (1.5). From the triangle inequality we have that for any $\vartheta \in \Theta$, $|A_{\vartheta, j, n}^{\circ}(\lambda)| \leq |A_{\vartheta, j, n}^{\circ}(\lambda) - A_{\vartheta}(\frac{j}{n}, \lambda)| + |A_{\vartheta}(\frac{j}{n}, \lambda)|$, $j = 1, \dots, n$. Using eq. (1.2) and the assumption that $A_{\vartheta}(j/n, \lambda)$ is uniformly bounded, we can refine this inequality. Indeed, for $n \geq 1$ we have that there is a generic constant such that

$$\sup_{i, \lambda} |A_{\vartheta, j, n}^{\circ}(\lambda)| \leq C. \quad (3.11)$$

Set $B_{\vartheta, j, k, n}^{\circ}(\lambda) = A_{\vartheta, j, n}^{\circ}(\lambda) \bar{A}_{\vartheta, k, n}^{\circ}(\lambda)$ and observe that

$$\begin{aligned} B_{\vartheta, j, k, n}^{\circ}(\lambda) &= \left[A_{\vartheta, j, n}^{\circ}(\lambda) - A_{\vartheta}\left(\frac{j}{n}, \lambda\right) \right] \bar{A}_{\vartheta, k, n}^{\circ}(\lambda) \\ &\quad + A_{\vartheta}\left(\frac{j}{n}, \lambda\right) \left[\bar{A}_{\vartheta, k, n}^{\circ}(\lambda) - \bar{A}_{\vartheta}\left(\frac{k}{n}, \lambda\right) \right] \\ &\quad + A_{\vartheta}\left(\frac{j}{n}, \lambda\right) \bar{A}_{\vartheta}\left(\frac{k}{n}, \lambda\right). \end{aligned}$$

This equality allows us to see that

$$\begin{aligned}
|B_{\vartheta_n, j, k, n}^\circ(\lambda) - B_{\vartheta_0, j, k, n}^\circ(\lambda)| &\leq \sup_{j, \lambda} \left| A_{\vartheta_n, j, n}^\circ(\lambda) - A_{\vartheta_n} \left(\frac{j}{n}, \lambda \right) \right| |A_{\vartheta_n, k, n}(-\lambda)| \\
&\quad + \sup_{k, n} \left| A_{\vartheta_0, j, n}^\circ(\lambda) - A_{\vartheta_0} \left(\frac{j}{n}, \lambda \right) \right| | - A_{\vartheta_0, k, n}(-\lambda) | \\
&\quad + |A_{\vartheta_n} \left(\frac{j}{n}, \lambda \right)| \sup_{k, n} \left| A_{\vartheta_n, k, n}^\circ(-\lambda) - A_{\vartheta_n} \left(\frac{k}{n}, -\lambda \right) \right| \\
&\quad + | - A_{\vartheta_0} \left(\frac{j}{n}, \lambda \right) | \sup_{k, n} \left| A_{\vartheta_0, k, n}^\circ(-\lambda) - A_{\vartheta_0} \left(\frac{k}{n}, -\lambda \right) \right| \\
&\quad + \left| A_{\vartheta_n} \left(\frac{j}{n}, \lambda \right) A_{\vartheta_n} \left(\frac{k}{n}, -\lambda \right) - A_{\vartheta_0} \left(\frac{j}{n}, \lambda \right) A_{\vartheta_0} \left(\frac{k}{n}, -\lambda \right) \right|.
\end{aligned}$$

From eqs. (1.2) and (3.11) we obtain that

$$|B_{\vartheta_n, j, k, n}^\circ(\lambda) - B_{\vartheta_0, j, k, n}^\circ(\lambda)| \leq \frac{C}{n} + \sup_{u, \lambda} \left| |A_{\vartheta_n}(u, \lambda)|^2 - |A_{\vartheta_0}(u, \lambda)|^2 \right|. \quad (3.12)$$

Set $B_{\vartheta_n, j, k, n}^*(\lambda) = B_{\vartheta_n, j, k, n}^\circ(\lambda) - B_{\vartheta_0, j, k, n}^\circ(\lambda)$,

$$\begin{aligned}
\|\Sigma_{\vartheta_n} - \Sigma_{\vartheta_0}\| &= \sup_{x \in \mathbb{C}^n} \frac{\left\| \left\{ \int_{-\pi}^{\pi} \exp\{i\lambda(j-k)\} B_{\vartheta_n, j, k, n}^*(\lambda) d\lambda \right\}_{j, k=1, \dots, n} x \right\|}{\|x\|} \\
&= 2\pi \sup_{u, \lambda} \left| f_{\vartheta_n}(u, \lambda) - f_{\vartheta_0}(u, \lambda) \right| + o(1) = o(1).
\end{aligned}$$

The latter follows from Assumption 3.1.2 and (3.12) above. \square

Lemma 3.16. *If $\vartheta_n \rightarrow \vartheta_0$ as $n \rightarrow \infty$, then $\nabla^2 \mathcal{L}_n(\vartheta_n) - \nabla^2 \mathcal{L}_n(\vartheta_0) \rightarrow 0$ in $\mathbf{P}_{\vartheta_n, n}$ -probability.*

Proof. Utilizing eq.(3.4), we have that for any $\vartheta \in \Theta$

$$\begin{aligned}
\nabla_{ij}^2 \mathcal{L}_n(\vartheta) &= -\frac{1}{2n} \text{tr} \left(\Sigma_{\vartheta}^{-1} C_{\vartheta}^{(i)} \Sigma_{\vartheta}^{-1} C_{\vartheta}^{(j)} \right) + \frac{1}{2n} \text{tr} \left(\Sigma_{\vartheta}^{-1} D_{\vartheta}^{(i,j)} \right) \\
&\quad + \frac{1}{2n} (X - \mu_{\vartheta})^{\top} \left(\nabla_{ij}^2 \Sigma_{\vartheta}^{-1} \right) (X - \mu_{\vartheta}) \\
&\quad + \frac{1}{n} (\nabla_i \mu_{\vartheta})^{\top} \Sigma_{\vartheta}^{-1} (\nabla_j \mu_{\vartheta}) \\
&\quad - \frac{1}{n} (\nabla_i \mu_{\vartheta})^{\top} \left(\nabla_j \Sigma_{\vartheta}^{-1} \right) (X - \mu_{\vartheta}) \\
&\quad - \frac{1}{n} (\nabla_j \mu_{\vartheta})^{\top} \left(\nabla_i \Sigma_{\vartheta}^{-1} \right) (X - \mu_{\vartheta}) \\
&\quad - \frac{1}{n} (\nabla_{ij}^2 \mu_{\vartheta})^{\top} \Sigma_{\vartheta}^{-1} (X - \mu_{\vartheta}).
\end{aligned} \tag{3.13}$$

Consider $\nabla_{ij}^2 \mathcal{L}_n(\vartheta_n) - \nabla_{ij}^2 \mathcal{L}_n(\vartheta_0) = I_n + II_n + III_n$, where

$$\begin{aligned}
I_n &= \frac{1}{2n} \text{tr} \left\{ \left(\Sigma_{\vartheta_0}^{-1} - \Sigma_{\vartheta_n}^{-1} \right) C_{\vartheta_0}^{(i)} \Sigma_{\vartheta_0}^{-1} C_{\vartheta_0}^{(j)} \right\} + \frac{1}{2n} \text{tr} \left\{ \Sigma_{\vartheta_n}^{-1} \left(C_{\vartheta_0}^{(i)} - C_{\vartheta_n}^{(i)} \right) \Sigma_{\vartheta_0}^{-1} C_{\vartheta_0}^{(j)} \right\} \\
&\quad + \frac{1}{2n} \text{tr} \left\{ \Sigma_{\vartheta_n}^{-1} C_{\vartheta_n}^{(i)} \left[\Sigma_{\vartheta_0}^{-1} C_{\vartheta_0}^{(j)} - \Sigma_{\vartheta_n}^{-1} C_{\vartheta_n}^{(j)} \right] \right\} + \frac{1}{2n} \text{tr} \left\{ \Sigma_{\vartheta_n}^{-1} D_{\vartheta_n}^{(i,j)} - \Sigma_{\vartheta_0}^{-1} D_{\vartheta_0}^{(i,j)} \right\},
\end{aligned} \tag{3.14}$$

$$\begin{aligned}
II_n &= \frac{1}{2n} X^{\top} \left[\left(\nabla_{ij}^2 \Sigma_{\vartheta_n}^{-1} \right) - \left(\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1} \right) \right] X \\
&\quad + \frac{1}{n} \left[\mu_{\vartheta_0}^{\top} \left(\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1} \right) - \mu_{\vartheta_n}^{\top} \left(\nabla_{ij}^2 \Sigma_{\vartheta_n}^{-1} \right) \right] X \\
&\quad + \frac{1}{n} \left[(\nabla_i \mu_{\vartheta_0})^{\top} \left(\nabla_j \Sigma_{\vartheta_0}^{-1} \right) - (\nabla_i \mu_{\vartheta_n})^{\top} \left(\nabla_j \Sigma_{\vartheta_n}^{-1} \right) \right] X \\
&\quad + \frac{1}{n} \left[\left(\nabla_{ij}^2 \mu_{\vartheta_0} \right)^{\top} \Sigma_{\vartheta_0}^{-1} - \left(\nabla_{ij}^2 \mu_{\vartheta_n} \right)^{\top} \Sigma_{\vartheta_n}^{-1} \right] X,
\end{aligned} \tag{3.15}$$

and

$$\begin{aligned}
III_n &= \frac{1}{2n} \left[\mu_{\vartheta_n}^{\top} \left(\nabla_{ij}^2 \Sigma_{\vartheta_n}^{-1} \right) \mu_{\vartheta_n} - \mu_{\vartheta_0}^{\top} \left(\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1} \right) \mu_{\vartheta_0} \right] \\
&\quad + \frac{1}{n} \left[(\nabla_i \mu_{\vartheta_n})^{\top} \left(\Sigma_{\vartheta_n}^{-1} \right) (\nabla_j \mu_{\vartheta_n}) - (\nabla_i \mu_{\vartheta_0})^{\top} \left(\Sigma_{\vartheta_0}^{-1} \right) (\nabla_j \mu_{\vartheta_0}) \right] \\
&\quad + \frac{1}{n} \left[(\nabla_i \mu_{\vartheta_n})^{\top} \left(\nabla_j \Sigma_{\vartheta_n}^{-1} \right) \mu_{\vartheta_n} - (\nabla_i \mu_{\vartheta_0})^{\top} \left(\nabla_j \Sigma_{\vartheta_0}^{-1} \right) \mu_{\vartheta_0} \right] \\
&\quad + \frac{1}{n} \left[\left(\nabla_{ij}^2 \mu_{\vartheta_n} \right)^{\top} \Sigma_{\vartheta_n}^{-1} \mu_{\vartheta_n} - \left(\nabla_{ij}^2 \mu_{\vartheta_0} \right)^{\top} \Sigma_{\vartheta_0}^{-1} \mu_{\vartheta_0} \right].
\end{aligned} \tag{3.16}$$

Let us focus on the first summand in the right-hand side of eq. (3.14). From Proposition V 1.8 and Theorem X 1.1 of Bhatia [5] we know that

$$\begin{aligned} \frac{1}{2n} \left| \text{tr} \left\{ \left(\Sigma_{\vartheta_0}^{-1} - \Sigma_{\vartheta_n}^{-1} \right) C_{\vartheta_0}^{(i)} \Sigma_{\vartheta_0}^{-1} C_{\vartheta_0}^{(j)} \right\} \right| &\leq \frac{1}{2} \|\Sigma_{\vartheta_0}^{-1} - \Sigma_{\vartheta_n}^{-1}\| \|C_{\vartheta_0}^{(i)}\| \|\Sigma_{\vartheta_0}^{-1}\| \|C_{\vartheta_0}^{(j)}\| \\ &\leq \frac{1}{2} \|\Sigma_{\vartheta_0}^{-1}\| \|\Sigma_{\vartheta_n} - \Sigma_{\vartheta_0}\| \|\Sigma_{\vartheta_n}^{-1}\| \|C_{\vartheta_0}^{(i)}\| \|\Sigma_{\vartheta_0}^{-1}\| \|C_{\vartheta_0}^{(j)}\|. \end{aligned} \quad (3.17)$$

Since $\|\Sigma_{\vartheta_n}^{-1}\|$, $\|C_{\vartheta_0}^{(i)}\|$, $\|\Sigma_{\vartheta_0}^{-1}\|$, and $\|C_{\vartheta_0}^{(j)}\|$ are uniformly bounded (by constants) for any i, j , cf. Lemma A.4 of Dahlhaus [19], and $\|\Sigma_{\vartheta_0} - \Sigma_{\vartheta_n}\| \rightarrow 0$, cf. Lemma 3.15, it follows that the left-hand side of eq. (3.17) tends to zero.

For the second summand in the right-hand side of eq. (3.14), observe that

$$\frac{1}{2n} \left| \text{tr} \left\{ \Sigma_{\vartheta_n}^{-1} \left(C_{\vartheta_0}^{(i)} - C_{\vartheta_n}^{(i)} \right) \Sigma_{\vartheta_0}^{-1} C_{\vartheta_0}^{(j)} \right\} \right| \leq \frac{1}{2} \|\Sigma_{\vartheta_n}^{-1}\| \|C_{\vartheta_0}^{(i)} - C_{\vartheta_n}^{(i)}\| \|\Sigma_{\vartheta_0}^{-1}\| \|C_{\vartheta_0}^{(j)}\|. \quad (3.18)$$

In analogy with the proof of Lemma 3.15, it can be established that $\|C_{\vartheta_n}^{(i)} - C_{\vartheta_0}^{(i)}\| \rightarrow 0$ provided $\vartheta_n \rightarrow \vartheta_0$. Therefore, the left-hand side of eq. 3.18 tends to zero. Since $\|D_{\vartheta}^{(i,j)}\|$ is uniformly bounded (by constants) for any i, j , cf. Lemma A.7 of [19], the remaining summands of eq. (3.14) can be handled similarly. This shows that $I_n \rightarrow 0$ as $n \rightarrow \infty$.

We now consider eq. (3.15). Utilizing that for any $\vartheta \in \Theta$, $\frac{\partial}{\partial \vartheta} \Sigma_{\vartheta}^{-1} = -\Sigma_{\vartheta}^{-1} \left(\frac{\partial}{\partial \vartheta} \Sigma_{\vartheta} \right) \Sigma_{\vartheta}^{-1}$ it can be seen that

$$\nabla_{ij}^2 \Sigma_{\vartheta}^{-1} = -\Sigma_{\vartheta}^{-1} D_{\vartheta}^{(i,j)} \Sigma_{\vartheta}^{-1} + 2 \Sigma_{\vartheta}^{-1} C_{\vartheta}^{(i)} \Sigma_{\vartheta}^{-1} C_{\vartheta}^{(j)} \Sigma_{\vartheta}^{-1}. \quad (3.19)$$

Consequently, for the first summand of eq. (3.15) we have

$$\begin{aligned} \left| \frac{1}{2n} X^{\top} \left[\left(\nabla_{ij}^2 \Sigma_{\vartheta_n}^{-1} \right) - \left(\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1} \right) \right] X \right| &\leq \frac{1}{2n} \|X\|_2^2 \left\| \left(\nabla_{ij}^2 \Sigma_{\vartheta_n}^{-1} \right) - \left(\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1} \right) \right\| \\ &\leq \frac{1}{2n} \|X\|_2^2 \left\| \Sigma_{\vartheta_n}^{-1} \right\| \left(\|D_{\vartheta_0}^{(i,j)}\| \|\Sigma_{\vartheta_0}^{-1} - \Sigma_{\vartheta_n}^{-1}\| \right. \\ &\quad \left. + \|D_{\vartheta_0}^{(i,j)} - D_{\vartheta_n}^{(i,j)}\| \|\Sigma_{\vartheta_n}^{-1}\| \right). \end{aligned} \quad (3.20)$$

Lemma 2.2 of Dahlhaus [19] shows that $1/n \|X\|_2^2$ is bounded in probability. Since $\|\Sigma_{\vartheta_0}\|$, $\|\Sigma_{\vartheta_n}\|$, and $\|D_{\vartheta_0}^{(i,j)}\|$ are uniformly bounded (by constants) and $\|\Sigma_{\vartheta_n} - \Sigma_{\vartheta_0}\| \rightarrow 0$ as $n \rightarrow \infty$,

∞ it follows that the first summand in the right-hand side of eq. (3.20) tends to zero. From Assumption 3.1.2 it follows that $D_{\vartheta}^{i,j}$ is a continuous function on ϑ , hence $\|D_{\vartheta_0}^{(i,j)} - D_{\vartheta_n}^{(i,j)}\| \rightarrow 0$ for any i, j . Hence, the second summand in the right-hand side of eq. (3.20) tends to zero. Consequently, the left-hand side of eq. (3.20) tends to zero.

The second summand in eq. (3.15) can be handled by utilizing the Cauchy-Schwarz inequality. Namely,

$$\begin{aligned} \left| \frac{1}{n} \left[\mu_{\vartheta_0}^\top (\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1}) - \mu_{\vartheta_n}^\top (\nabla_{ij}^2 \Sigma_{\vartheta_n}^{-1}) \right] X \right| &\leq \frac{1}{n} \left(\|\mu_{\vartheta_0} - \mu_{\vartheta_n}\|_2^2 \|X\|_2^2 \right)^{1/2} \|\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1}\| \\ &+ \frac{1}{n} \left(\|\mu_{\vartheta_n}\|_2^2 \|X\|_2^2 \right)^{1/2} \left\| (\nabla_{ij}^2 \Sigma_{\vartheta_n}^{-1}) - (\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1}) \right\|. \end{aligned} \quad (3.21)$$

Assumption 3.1.3 yields that $\|\mu_{\vartheta_0} - \mu_{\vartheta_n}\|_2^2 \rightarrow 0$ provided $\vartheta_n \rightarrow \vartheta_0$. Similarly as above, it can be shown that $\left\| (\nabla_{ij}^2 \Sigma_{\vartheta_n}^{-1}) - (\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1}) \right\| \rightarrow 0$ as $n \rightarrow \infty$, for any i, j ; cf. eq. (3.19). The remaining terms on the right-hand side of eq. (3.21) are bounded in probability and consequently, the left-hand side of eq. (3.21) tends to zero. Since $\|\nabla \mu_{\vartheta}\|_2^2$ is bounded (for any $\vartheta \in \Theta$) and $\|\nabla \mu_{\vartheta_n} - \nabla \mu_{\vartheta_0}\|_2^2 \rightarrow 0$; cf. Assumption 3.1.3, the remaining summands of eq. (3.15) can be handled similarly. This shows that $II_n \rightarrow 0$ as $n \rightarrow \infty$.

Finally, note that terms in (3.16) have the form

$$\frac{1}{n} \mathbf{a}_{\vartheta}^\top M_{\vartheta} \mathbf{b}_{\vartheta}$$

where $\|M_{\vartheta_0} - M_{\vartheta_n}\| \rightarrow 0$, $\frac{1}{n} \|\mathbf{a}_{\vartheta_0} - \mathbf{a}_{\vartheta_n}\|_2^2 \rightarrow 0$ provided $\vartheta_n \rightarrow \vartheta_0$, and $\|\mathbf{b}_{\vartheta}\|$ is bounded. These terms can be handled in analogy with the treatment for I_n and II_n . This shows that $III_n \rightarrow 0$ as $n \rightarrow \infty$. This completes the proof. \square

Lemma 3.17. *If $\vartheta_n \rightarrow \vartheta_0$ as $n \rightarrow \infty$, then $\nabla^2 \mathcal{L}_n(\vartheta_0) \rightarrow \Gamma_{\vartheta_0}$ in $\mathbf{P}_{\vartheta_n, n}$ -probability.*

Proof. By Chebyshev's inequality, it suffices to show that $\mathbb{E}_{\vartheta_n}[\nabla_{ij}^2 \mathcal{L}_n(\vartheta_0)] \rightarrow (\Gamma_{\vartheta_0})_{ij}$ and $\mathbb{V}_{\vartheta_n}[\nabla^2 \mathcal{L}_n(\vartheta_0)_{ij}] = o(n^{-1})$ for $i, j = 1, \dots, d$. We begin with the calculations for the

expectation. Set $Y_n = \frac{1}{2n} (X - \mu_{\vartheta_0})^\top \left(\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1} \right) (X - \mu_{\vartheta_0})$ and use (3.13) to see that for any $i, j = 1, \dots, d$:

$$\begin{aligned}
\mathbb{E}_{\vartheta_n} \left[\nabla_{ij}^2 \mathcal{L}_n(\vartheta_0) \right] &= -\frac{1}{2n} \text{tr} \left(\Sigma_{\vartheta_0}^{-1} C_{\vartheta_0}^{(i)} \Sigma_{\vartheta_0}^{-1} C_{\vartheta_0}^{(j)} \right) + \frac{1}{2n} \text{tr} \left(\Sigma_{\vartheta_0}^{-1} D_{\vartheta_0}^{(i,j)} \right) + \mathbb{E}_{\vartheta_n} [Y_n] \\
&\quad + \frac{1}{n} (\nabla_i \mu_{\vartheta_0})^\top \Sigma_{\vartheta_0}^{-1} (\nabla_j \mu_{\vartheta_0}) \\
&\quad - \frac{1}{n} (\nabla_i \mu_{\vartheta_0})^\top \left(\nabla_j \Sigma_{\vartheta_0}^{-1} \right) (\mu_{\vartheta_n} - \mu_{\vartheta_0}) \\
&\quad - \frac{1}{n} (\nabla_j \mu_{\vartheta_0})^\top \left(\nabla_i \Sigma_{\vartheta_0}^{-1} \right) (\mu_{\vartheta_n} - \mu_{\vartheta_0}) \\
&\quad - \frac{1}{n} (\nabla_{ij}^2 \mu_{\vartheta_0})^\top \Sigma_{\vartheta_0}^{-1} (\mu_{\vartheta_n} - \mu_{\vartheta_0}). \tag{3.22}
\end{aligned}$$

Recall that for any random vector $z \sim \mathcal{N}(0, \Sigma)$, $\mathbb{E}[z^\top A z] = \text{tr}(A \Sigma)$ and also $\mathbb{V}[z^\top A z] = 2 \text{tr}(A \Sigma A \Sigma)$, it follows that if $X \sim \mathcal{N}(\mu_{\vartheta_n}, \Sigma_{\vartheta_n})$ then

$$\begin{aligned}
2n \mathbb{E}_{\vartheta_n} [Y_n] &= \text{tr} \left\{ \left(\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1} \right) \Sigma_{\vartheta_n} \right\} \\
&\quad + (\mu_{\vartheta_n} - \mu_{\vartheta_0})^\top \left(\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1} \right) (\mu_{\vartheta_n} - \mu_{\vartheta_0}) \tag{3.23}
\end{aligned}$$

$$\begin{aligned}
4n^2 \mathbb{V}_{\vartheta_n} [Y_n] &= 2 \text{tr} \left\{ \left(\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1} \right) \Sigma_{\vartheta_n} \left(\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1} \right) \Sigma_{\vartheta_n} \right\} \\
&\quad + 4 (\mu_{\vartheta_n} - \mu_{\vartheta_0})^\top \left(\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1} \right) \Sigma_{\vartheta_n} \left(\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1} \right) (\mu_{\vartheta_n} - \mu_{\vartheta_0}). \tag{3.24}
\end{aligned}$$

By properly substituting eq. (3.23) into eq. (3.22) we obtain:

$$\begin{aligned}
\mathbb{E}_{\vartheta_n} \left[\nabla_{ij}^2 \mathcal{L}_n(\vartheta_0) \right] &= -\frac{1}{2n} \text{tr} \left(\Sigma_{\vartheta_0}^{-1} C_{\vartheta_0}^{(i)} \Sigma_{\vartheta_0}^{-1} C_{\vartheta_0}^{(j)} \right) + \frac{1}{2n} \text{tr} \left(\Sigma_{\vartheta_0}^{-1} D_{\vartheta_0}^{(i,j)} \right) \\
&\quad - \frac{1}{2n} \text{tr} \left\{ \Sigma_{\vartheta_n} \Sigma_{\vartheta_0}^{-1} D_{\vartheta_0}^{(i,j)} \Sigma_{\vartheta_0}^{-1} \right\} \\
&\quad + \frac{1}{n} \text{tr} \left\{ \Sigma_{\vartheta_n} \Sigma_{\vartheta_0}^{-1} C_{\vartheta_0}^{(i)} \Sigma_{\vartheta_0}^{-1} C_{\vartheta_0}^{(j)} \Sigma_{\vartheta_0}^{-1} \right\} \\
&\quad + \frac{1}{2n} (\mu_{\vartheta_n} - \mu_{\vartheta_0})^\top \left(\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1} \right) (\mu_{\vartheta_n} - \mu_{\vartheta_0}) \\
&\quad + \frac{1}{n} (\nabla_i \mu_{\vartheta_0})^\top \Sigma_{\vartheta_0}^{-1} (\nabla_j \mu_{\vartheta_0}) \\
&\quad - \frac{1}{n} (\nabla_i \mu_{\vartheta_0})^\top \left(\nabla_j \Sigma_{\vartheta_0}^{-1} \right) (\mu_{\vartheta_n} - \mu_{\vartheta_0}) \\
&\quad - \frac{1}{n} (\nabla_j \mu_{\vartheta_0})^\top \left(\nabla_i \Sigma_{\vartheta_0}^{-1} \right) (\mu_{\vartheta_n} - \mu_{\vartheta_0}) \\
&\quad - \frac{1}{n} (\nabla_{ij}^2 \mu_{\vartheta_0})^\top \Sigma_{\vartheta_0}^{-1} (\mu_{\vartheta_n} - \mu_{\vartheta_0}), \tag{3.25}
\end{aligned}$$

From Lemma 4.8 of Dahlhaus [19] we obtain that the summands in the right-hand side of (3.25) tend to

$$\begin{aligned}
&-\frac{1}{4\pi} \int_{\mathbb{T}} \frac{\left(\frac{\partial}{\partial \vartheta_i} f_{\vartheta_0} \right) \left(\frac{\partial}{\partial \vartheta_j} f_{\vartheta_0} \right)}{f_{\vartheta_0}^2} + \frac{1}{4\pi} \int_{\mathbb{T}} \left(\frac{\frac{\partial}{\partial \vartheta_i} \frac{\partial}{\partial \vartheta_j} f_{\vartheta_0}}{f_{\vartheta_0}} \right) \\
&-\frac{1}{4\pi} \int_{\mathbb{T}} f_{\vartheta_n} \left(\frac{\frac{\partial}{\partial \vartheta_i} \frac{\partial}{\partial \vartheta_j} f_{\vartheta_0}}{f_{\vartheta_0}^2} \right) + \frac{1}{2\pi} \int_{\mathbb{T}} f_{\vartheta_n} \frac{\left(\frac{\partial}{\partial \vartheta_i} f_{\vartheta_0} \right) \left(\frac{\partial}{\partial \vartheta_j} f_{\vartheta_0} \right)}{f_{\vartheta_0}^3}
\end{aligned}$$

at the rate $n^{-2/3} \log^4(n)$.

Because of the smoothness impose on f_{ϑ} , the terms above are indeed equal to

$$\frac{1}{4\pi} \int_{\mathbb{T}} \frac{\left(\frac{\partial}{\partial \vartheta_i} f_{\vartheta_0} \right) \left(\frac{\partial}{\partial \vartheta_j} f_{\vartheta_0} \right)}{f_{\vartheta_0}^2} = (\Gamma_{\vartheta_0})_{ij}.$$

This shows that the first of our claims holds true.

We now move onto the corresponding calculations for the variance. Observe that

combining (3.13), (3.24) and (3.19)

$$\begin{aligned} \mathbf{V}_{\vartheta_n} [\nabla_{ij}^2 \mathcal{L}_n(\vartheta_0)] &= \frac{1}{2n^2} \text{tr} \left\{ (\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1}) \Sigma_{\vartheta_n} (\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1}) \Sigma_{\vartheta_n} \right\} \\ &\quad + \frac{1}{n^2} (\nabla_i \mu_{\vartheta_0})^\top (\nabla_j \Sigma_{\vartheta_0}^{-1}) \Sigma_{\vartheta_n} (\nabla_j \Sigma_{\vartheta_0}^{-1})^\top \nabla_i \mu_{\vartheta_0} \\ &\quad + \frac{1}{n^2} (\nabla_j \mu_{\vartheta_0})^\top (\nabla_i \Sigma_{\vartheta_0}^{-1}) \Sigma_{\vartheta_n} (\nabla_i \Sigma_{\vartheta_0}^{-1})^\top \nabla_j \mu_{\vartheta_0} \\ &\quad + \frac{1}{n^2} (\nabla_{ij}^2 \mu_{\vartheta_0})^\top \Sigma_{\vartheta_0}^{-1} \Sigma_{\vartheta_n} (\Sigma_{\vartheta_0}^{-1})^\top (\nabla_{ij}^2 \mu_{\vartheta_0}). \end{aligned}$$

Moreover, combining eq. (3.24) and eq. (3.19), the expression above becomes

$$\begin{aligned} \mathbf{V}_{\vartheta_n} [\nabla_{ij}^2 \mathcal{L}_n(\vartheta_0)] &= \frac{1}{2n^2} \text{tr} \left\{ (\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1}) \Sigma_{\vartheta_n} (\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1}) \Sigma_{\vartheta_n} \right\} \\ &\quad + \frac{1}{n^2} (\mu_{\vartheta_n} - \mu_{\vartheta_0})^\top (\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1}) \Sigma_{\vartheta_n} (\nabla_{ij}^2 \Sigma_{\vartheta_0}^{-1}) (\mu_{\vartheta_n} - \mu_{\vartheta_0}) \\ &\quad + \frac{1}{n^2} (\nabla_i \mu_{\vartheta_0})^\top (\nabla_j \Sigma_{\vartheta_0}^{-1}) \Sigma_{\vartheta_n} (\nabla_j \Sigma_{\vartheta_0}^{-1})^\top \nabla_i \mu_{\vartheta_0} \\ &\quad + \frac{1}{n^2} (\nabla_j \mu_{\vartheta_0})^\top (\nabla_i \Sigma_{\vartheta_0}^{-1}) \Sigma_{\vartheta_n} (\nabla_i \Sigma_{\vartheta_0}^{-1})^\top \nabla_j \mu_{\vartheta_0} \\ &\quad + \frac{1}{n^2} (\nabla_{ij}^2 \mu_{\vartheta_0})^\top \Sigma_{\vartheta_0}^{-1} \Sigma_{\vartheta_n} (\Sigma_{\vartheta_0}^{-1})^\top (\nabla_{ij}^2 \mu_{\vartheta_0}). \end{aligned} \quad (3.26)$$

Once again apply Lemma 4.8 in [19] to eq. (3.26) with $\vartheta = \vartheta_0$ to obtain that the variance $\mathbf{V}_{\vartheta_n} [\nabla_{ij}^2 \mathcal{L}_n(\vartheta_0)] = o(n^{-1})$. This completes the proof. \square

Lemma 3.18. *Let $K \in \mathcal{K}_\Theta$ and suppose that Assumption 3.1 holds. Then, for any sequence of points $\vartheta_n \in K$ and any sequence of points $u_n \rightarrow u \in \mathbb{R}^d$, the following weak convergence*

$$\mathfrak{L}(\langle u_n, S_n(\vartheta_n) \rangle \mid \mathbf{P}_{\vartheta_n, n}) \Rightarrow \mathfrak{L}(\langle u, Z \rangle), \quad Z \sim \mathcal{N}(0, I_d), \quad (3.27)$$

where $S_n(\vartheta_n) = -\sqrt{n} \Gamma_{\vartheta_n}^{-1} \nabla \mathcal{L}_n(\vartheta_n)$, is valid. Recall that Γ_{ϑ_0} is defined in eq. (3.5).

Proof. Since K is compact, it suffices to prove this lemma for any convergent sequence $\{\vartheta_n\}$. To ease the notation, write $d_n = \langle u_n, S_n(\vartheta_n) \rangle$. From eq. (3.4) and using the Assumption 3.1.5, it follows that $\mathbf{E}_{\vartheta_n}[\sqrt{n} \nabla \mathcal{L}_n(\vartheta_n)] \rightarrow 0$. Consequently, $\mathbf{E}_{\vartheta_n}[d_n] \rightarrow 0$.

Now observe that

$$\mathbf{V}_{\vartheta_n}[d_n] = \mathbf{u}_n^\top \Gamma_{\vartheta_n}^{-1/2} \mathbf{V}[\sqrt{n} \nabla \mathcal{L}_n(\vartheta_n)] \Gamma_{\vartheta_n}^{-1/2} \mathbf{u}_n$$

Since $\mathbf{u}_n \rightarrow \mathbf{u}$, in order to complete the proof it suffices to show that:

$$n \left(\mathbf{V}_{\vartheta_n} \left[\nabla_{i,j} \mathcal{L}_n(\vartheta_n) \right] - (\Gamma_{\vartheta_n})_{i,j} \right) \rightarrow 0,$$

for $i, j = 1, \dots, d$.

To this end, we utilize the method of cumulants. Firstly, note that for any $i, j \in \{1, \dots, d\}$

$$n \left| \text{cov} \left(\nabla_i \mathcal{L}_n(\vartheta_n), \nabla_j \mathcal{L}_n(\vartheta_n) \right) - (\Gamma_{\vartheta_n})_{i,j} \right| = \left| \frac{1}{2n} \text{tr} \left\{ \Sigma_{\vartheta_n}^{-1} C_{\vartheta_n}^{(i)} \Sigma_{\vartheta_n}^{-1} C_{\vartheta_n}^{(j)} \right\} - (\Gamma_{\vartheta_n})_{i,j} \right| \rightarrow 0,$$

cf. Lemma 4.8 of Dahlhaus [19]. Replacing ϑ_0 by ϑ_n , the rest of this proof is analogous to that of Theorem 2.4(iii) of [19]. □

CHAPTER 4

GLOBAL PARAMETRIC ASYMPTOTIC EQUIVALENCE OF LOCALLY STATIONARY PROCESSES

Consider the statistical experiment given by observations from a locally stationary Gaussian process with zero mean and time-varying spectral density f_ϑ , $\vartheta \in \Theta \subset \mathbb{R}^d$. The global asymptotic equivalence in LeCam's sense of this experiment and a white noise experiment with drift $\log f_\vartheta$ is established. This result is obtained by an application of LeCam's connection theorem.

4.1 Introduction and main result

Let $X := (X_{k,n})_{1 \leq k \leq n}$ be a locally stationary Gaussian process (lsGp) with zero trend function, transfer function A° and approximating function A . A proper definition of these terms can be found in the introductory chapter of this work, cf. Definition 1.1. We have previously explained the asymptotic relation between the covariance matrix of X and the time-varying spectral density (tvsd) function $f = (2\pi)^{-1}|A|^2$. Recall for instance that f is the limit, in $\mathbb{L}_2(-\pi, \pi)$, of the Wigner-Ville spectrum, cf. Ch. 1. Gaussianity and the assumption of zero trend makes the statistical inference about X exclusively dependable on proper estimates of f .

In this chapter we study the properties of f_ϑ , an estimate of f which is parametrized by a vector $\vartheta \in \Theta \subset \mathbb{R}^d$. This assumption is not only natural in statistical modeling but is also in line with the applications of models based on locally stationary Gaussian processes found in the literature. For instance, recently, models based on these processes have been reported in areas such as seismology, medicine and computer science; cf. Beran [4], Eckley et al. [25] and Maharaj and Alonso [37].

We are interested in the problem of estimating the time-varying spectral density function $f_\vartheta = (2\pi)^{-1}|A_\vartheta|^2$ from the point of view of LeCam's theory of statistical experiments. Especially, we show that the decision theoretic properties of estimators of f_ϑ when using a sample from a locally stationary Gaussian process are asymptotically equivalent to those of estimators of $\log f_\vartheta$ in the white noise problem $dZ_{(u,t)} = \log f_\vartheta(u, t) du dt + 2\sqrt{\pi}n^{-1/2} d\mathbf{W}_{(u,t)}$, where \mathbf{W} is a bidimensional Brownian motion on $\mathbb{T} = [0, 1] \times [-\pi, \pi]$.

We now properly state the main result of this chapter. First, some notation is in order. Define the parameter set

$$\mathfrak{F}_\Theta := \left\{ f_\vartheta = \frac{1}{2\pi}|A_\vartheta|^2, \vartheta \in \Theta \subset \mathbb{R}^d : A_\vartheta \text{ satisfies Assumption 3.1} \right\}. \quad (4.1)$$

Let Δ be the LeCam's pseudodistance between experiments having the same parameter space. Two sequences of statistical experiments \mathbb{F}_n and \mathbb{G}_n are said to be asymptotically equivalent in the sense of LeCam if $\Delta(\mathbb{F}_n, \mathbb{G}_n) \rightarrow 0$ as $n \rightarrow \infty$. To ease the reading, a formal definition of this type of convergence is given in Section 4.2.

Theorem 4.1. *The experiments given by observations*

$$X_{1,n}, \dots, X_{n,n} \text{ from a lsGp with zero trend function and tvsd } f_\vartheta, \text{ and} \quad (4.2)$$

$$dZ_{(u,\lambda)} = \log f_\vartheta(u, \lambda) du d\lambda + \frac{2\sqrt{\pi}}{\sqrt{n}} d\mathbf{W}_{(u,\lambda)}, \quad (u, \lambda) \in \mathbb{T}, \quad (4.3)$$

are asymptotically equivalent in the sense of LeCam for any time-varying spectral density function $f_\vartheta \in \mathfrak{F}_\Theta$.

Remark 4.2. *The assumption that the trend function is identically zero in Theorem 4.1 is quite unsatisfactory. Indeed, it is expected that the mean of a locally stationary process varies, at least smoothly, even in a very small period of time. To carry out statistical inference for stationary processes, however, it has been customary to assume that the trend is negligible, cf. Walker [48], Davies [23] and Golubev et al. [29].*

Theorem 4.1 provides a clear picture of the time-varying spectral density estimation problem in the case in which the parameter space is finite dimensional. In this manner, our result also precludes what can be expected in a nonparametric estimation framework.

This theorem is obtained as an application of a general result known as the *connection theorem* which is due to Lucien LeCam [33]. LeCam's connection theorem as well as the essentials of LeCam's theory of statistical experiments are introduced in Section 4.2. Roughly speaking, this method requires that two conditions be fulfilled: (1) the asymptotic equivalence must hold when the parameter space is restricted to specific subspaces, and (2) there must exist suitable estimators of the parameter of interest on each of the global experiments under consideration. Sections 4.3 and 4.4 deal with (1) and (2), respectively.

In the concluding Section 4.5, we introduce λ -convergence of statistical experiments. This concept is due to Shiryaev and Spokoiny [43] and it is of interest on its own, particularly because in combination with uniform local asymptotic normality (ULAN), λ -convergence implies the convergence of local experiments in the sense of LeCam.

4.2 From local to global equivalence: LeCam's connection theorem

One of the basic features of LeCam's theory of statistical experiments is that we can approximate general experiments by simple ones. Here, the approximation methods are based on decision theoretic principles. This latter characteristic makes that LeCam's theory be suitable to perform asymptotic statistical inference in an abstract manner. We begin this section with the essentials of LeCam theory.

Let $\mathbb{E}_{1,n} = (\Omega_{1,n}, \mathcal{A}_{1,n}, \{\mathbf{P}_{\vartheta}^n : \vartheta \in \Theta\})$ and $\mathbb{E}_{2,n} = (\Omega_{2,n}, \mathcal{A}_{2,n}, \{\mathbf{Q}_{\vartheta}^n : \vartheta \in \Theta\})$ be two

sequences of statistical experiments. Here $(\Omega_{i,n}, \mathcal{A}_{i,n})$, $i = 1, 2$, are measurable spaces on which the families of probabilities $\{\mathbf{P}_\vartheta^n; \vartheta \in \Theta\}$ and $\{\mathbf{Q}_\vartheta^n; \vartheta \in \Theta\}$ are defined. Unless stated otherwise, throughout this section Θ denotes an arbitrary parameter space.

By definition, LeCam's Δ -pseudodistance between $\mathbb{E}_{1,n}$ and $\mathbb{E}_{2,n}$ is given by

$$\Delta(\mathbb{E}_{1,n}, \mathbb{E}_{2,n}; \Theta) = \max(\delta(\mathbb{E}_{1,n}, \mathbb{E}_{2,n}; \Theta), \delta(\mathbb{E}_{2,n}, \mathbb{E}_{1,n}; \Theta)),$$

where $\delta(\mathbb{E}_{1,n}, \mathbb{E}_{2,n}; \Theta)$ denotes the deficiency of $\mathbb{E}_{1,n}$ with respect to $\mathbb{E}_{2,n}$ and is given by

$$\delta(\mathbb{E}_{1,n}, \mathbb{E}_{2,n}; \Theta) = \inf_T \sup_{\vartheta \in \Theta} \|T \mathbf{P}_\vartheta^n - \mathbf{Q}_\vartheta^n\|_{\text{TV}}.$$

Here $\|\cdot\|_{\text{TV}}$ denotes the total variation distance between probability measures and the infimum is taken over all the transition functions, T , from $(\Omega_{1,n}, \mathcal{A}_{1,n})$ to $(\Omega_{2,n}, \mathcal{A}_{2,n})$; cf. Strasser [44] or Section 9 of Nussbaum [40] for more details about these terms.

Although the triangle inequality is held, i.e., $\Delta(\mathbb{E}, \mathbb{G}) \leq \Delta(\mathbb{E}, \mathbb{F}) + \Delta(\mathbb{F}, \mathbb{G})$ provided \mathbb{E} , \mathbb{F} and \mathbb{G} have the same parameter space, properly speaking Δ is a pseudo distance because $\Delta(\mathbb{E}, \mathbb{F}) = 0$ does not imply that $\mathbb{E} \equiv \mathbb{F}$.

The experiments $\mathbb{E}_{1,n}$ and $\mathbb{E}_{2,n}$ are said to be asymptotically equivalent in LeCam's sense if $\Delta(\mathbb{E}_{1,n}, \mathbb{E}_{2,n}; \Theta) \rightarrow 0$ as $n \rightarrow \infty$. This can be interpreted as follows: for any statistical procedure, say ρ_1 , defined on $\mathbb{E}_{1,n}(\mathbb{E}_{2,n})$, and any $\varepsilon > 0$, there is a counterpart, say ρ_2 , defined on $\mathbb{E}_{2,n}(\mathbb{E}_{1,n})$ such that for every bounded loss function ℓ , the risk of ρ_2 is ε -close to the risk of ρ_1 provided $n \rightarrow \infty$. In other words, an investigation in $\mathbb{E}_{1,n}$ would automatically yields analogous results in $\mathbb{E}_{2,n}$ and viceversa. Proofs of this description of the Δ -distance can be found in Brown and Low [9] or in LeCam and Yang [36].

Classical nonparametric estimation problems have been studied under the viewpoint of LeCam theory. For instance, it is known that asymptotically, regression and density estimation both can be seen as white noise problems, cf. [9] and [40], respectively.

More recently, for stationary time series, Golubev et al. [29] established that nonparametric spectral density estimation is asymptotically equivalent to a white noise problem; asymptotic equivalence for nonparametric regression problems with non regular errors have been studied by Meister and Reiss [38].

For statistical experiments whose parameter space Θ has metric-dimensionality restrictions, e.g., if Θ is a subset of \mathbb{R}^d , LeCam introduced a procedure to deal with asymptotic equivalence: the connection theorem. Roughly speaking, this method requires that two conditions be fulfilled: (1) the asymptotic equivalence must hold when the parameter space is restricted to subspaces of radius $b < \infty$, and (2) there must exist a class of estimators of the parameter of interest on each of the global experiments under consideration whose rate of convergence is compatible with b . Now, we formally introduce LeCam's connection theorem and present some of its applications.

Let $\mathbb{E}_{1,n}$ and $\mathbb{E}_{2,n}$ be two sequences of experiments having the same parameter space Θ . Suppose that Θ is metrized by m . Let $\Theta' \subseteq \Theta$, the diameter of Θ' is defined as $\text{diam}(\Theta') := \sup\{m(\vartheta, \nu) : \vartheta, \nu \in \Theta'\}$. Define the restricted experiments $\mathbb{E}_{1,n}(\Theta') := \{\mathbf{P}_{\vartheta}^n : \vartheta \in \Theta'\}$, and $\mathbb{E}_{2,n}(\Theta') := \{\mathbf{Q}_{\vartheta}^n : \vartheta \in \Theta'\}$. If for given sequences of numbers $\{a_n\}$ and $\{b_n\}$, $0 < a_n \leq b_n$ we have that:

1. For any $\varepsilon > 0$, and for any subset Θ' whose diameter is less than $3b_n$,

$$\delta(\mathbb{E}_{1,n}(\Theta'), \mathbb{E}_{2,n}(\Theta'); \Theta') < \frac{\varepsilon}{2},$$

and

2. If within $\mathbb{E}_{1,n}$, there exists a sequence of estimates $\widehat{\vartheta}_n$ such that for any $\varepsilon > 0$ and for each $\vartheta \in \Theta$,

$$\mathbf{P}_{\vartheta}^n\{m(\widehat{\vartheta}_n, \vartheta) > a_n\} < \frac{\varepsilon}{2}.$$

Then, $\delta(\mathbb{E}_{1,n}, \mathbb{E}_{2,n}; \Theta) < \varepsilon + \frac{a_n}{b_n} C$, where C is the maximum number of sets needed to cover subsets of diameter $4a_n + 2b_n$; cf. Theorem 1 of LeCam [34]. Note that if $\frac{a_n}{b_n} C \rightarrow 0$ as $n \rightarrow \infty$ the interpolation formula provides a method which might prove useful to show asymptotic equivalence between experiments when the parameter space meets some finite metric-dimensionality restrictions.

As an application, we now present a proof of the asymptotic equivalence of nonparametric density estimation and white noise problem when the set of densities has finite Hellinger metric dimension.

Example 4.3. Let Θ be a set of densities on $[0, 1]$. For $\vartheta_1, \vartheta_2 \in \Theta$ define their square Hellinger distance by

$$h^2(\vartheta_1, \vartheta_2) = \int_0^1 \left(\sqrt{\vartheta_1(x)} - \sqrt{\vartheta_2(x)} \right)^2 dx,$$

and for $n \in \mathbb{N}^+$, let Θ be metrized by the Hellinger metric, $\mathfrak{m}_n(\vartheta_1, \vartheta_2) = n h^2(\vartheta_1, \vartheta_2)$.

A set Θ^* in $\mathbb{L}_2(0, 1)$ is said to have finite metric dimension if there is a number D such that every subset of Θ^* which can be covered by an ε -ball, can be covered by no more than 2^D balls of radius $\varepsilon/2$. The set of densities ϑ ($\in \Theta$) has this property in Hellinger metric if the corresponding set of $\vartheta^{1/2}$ functions has it in $\mathbb{L}_2(0, 1)$.

Suppose that Θ has finite dimension in the Hellinger metric \mathfrak{m}_n and fulfill a further regularity condition. Let \mathbf{B} be a Brownian motion on $[0, 1]$. Then, the experiments given by observations

$$y_i, \quad i = 1, \dots, n \quad \text{i.i.d. with density } \vartheta,$$

$$d\vartheta(t) = \vartheta^{1/2}(t) dt + \frac{1}{2} n^{-1/2} d\mathbf{B}(t), \quad t \in [0, 1],$$

are asymptotically equivalent, where $\vartheta \in \Theta$. We can prove this result as follows.

Let $\mathbb{E}_{1,n}$ and $\mathbb{E}_{2,n}$, respectively, denote the experiments just defined. Let Δ denote the LeCam's pseudodistance. Then observe that

1. Let $\Theta_b(\vartheta) = \{\nu \in \Theta : m_n(\vartheta, \nu) \leq b\}$. From LeCam [35] we know that there exists a sequence $b_n \rightarrow \infty$ such that for n large enough,

$$\sup_{\vartheta \in \Theta} \Delta(\mathbb{E}_{1,n}, \mathbb{E}_{2,n}; \Theta_{b_n}(\vartheta)) \leq \varepsilon/2.$$

2. From Birgé [3] we know that in both experiments, $\mathbb{E}_{1,n}$ and $\mathbb{E}_{2,n}$, there are estimates $\widehat{\vartheta}_n$ and a number $\tau < \infty$ such that for n large enough,

$$\sup_{\vartheta} \mathbf{P}_{\vartheta} \left\{ m_n(\widehat{\vartheta}_n, \vartheta) \geq \tau \right\} \leq \varepsilon/2.$$

Consequently, $\Delta(\mathbb{E}_{1,n}, \mathbb{E}_{2,n}; \Theta) \leq \varepsilon + \frac{\tau}{b_n} C$, where $C < \infty$. Letting $n \rightarrow \infty$ the claim now follows.

We conclude this section with another application of LeCam's connection theorem: we utilize it to prove the validity of the main result of this chapter.

Proof of Theorem 4.1. Let Θ be an open subset of \mathbb{R}^d . For any $n \in \mathbb{N}^+$ let Θ be metrized by $m_n(\vartheta_1, \vartheta_2) = n\|\vartheta_1 - \vartheta_2\|_2^2$, $\vartheta_1, \vartheta_2 \in \mathbb{R}^d$. Let \mathbb{F}_n , \mathbb{G}_n , and \mathbb{E} be the experiments defined by eqs. (4.4), (4.5), and (4.6), respectively.

Let $K_b(\vartheta) = \{\nu \in \Theta : m_n(\vartheta, \nu) \leq b\}$ be a closed subset of Θ . $K_b(\vartheta)$ is a compact set since its diameter is finite. We can assume that we are given a sequence b_n such that $\sup_{\vartheta \in \Theta} \text{diam}(K_{b_n}(\vartheta)) < b_0 < \infty$. Proposition 4.4 yields that even for b_n large enough, $\Delta(\mathbb{F}_n, \mathbb{E}; K_{b_n}(\vartheta)) \leq \varepsilon$, and $\Delta(\mathbb{G}_n, \mathbb{E}; K_{b_n}(\vartheta)) \leq \varepsilon$. It is immediate from the triangle inequality that $\Delta(\mathbb{F}_n, \mathbb{G}_n; K_{b_n}(\vartheta)) \leq 2\varepsilon$.

From Corollary 3.7 and Proposition 4.12 we know that in both experiments, \mathbb{F}_n and \mathbb{G}_n , there are estimates $\widehat{\vartheta}_n$ and a number $\tau < \infty$ such that $\sup_{\vartheta} \mathbf{P}_{\vartheta, n} \{ m_n(\widehat{\vartheta}_n, \vartheta) > \tau \} \leq \varepsilon$.

Then, $\Delta(\mathbb{F}_n, \mathbb{G}_n; \Theta) \leq 3\varepsilon + \frac{\tau}{b_n} C$, where $C < \infty$. Let n tend to infinity to complete the proof. \square

4.3 Local asymptotic equivalence

The class of locally stationary Gaussian process considered in this chapter can be think of as a realization of an n -dimensional Gaussian random variable with zero mean and covariance matrix $\Sigma_\vartheta = \Sigma_n(A_\vartheta^\circ, A_\vartheta^\circ)$ where Σ_n is defined in (1.5). Thus it is notationally convenient to denote the statistical experiment introduced in Theorem 4.1, eq. (4.2), as

$$\mathbb{F}_n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\mathbf{P}_{\vartheta,n}; \vartheta \in \Theta\}), \quad (4.4)$$

where $\mathbf{P}_{\vartheta,n} = \mathcal{N}_n(0, \Sigma_\vartheta)$, $\vartheta \in \Theta \subset \mathbb{R}^d$.

Let \mathbf{W} be a Brownian motion defined on $\mathbb{T} = [0, 1] \times [-\pi, \pi]$, and for $f_\vartheta \in \mathcal{F}_\Theta$ consider an observed process

$$z_{(u,\lambda)} = \int_0^u \int_{-\pi}^\lambda \log f_\vartheta(s, t) ds dt + \frac{2\sqrt{\pi}}{\sqrt{n}} \mathbf{W}_{(u,\lambda)}, \quad (u, \lambda) \in \mathbb{T}.$$

Let $\mathbf{Q}_{\vartheta,n}$ be the distribution of this process on the space of continuous functions on \mathbb{T} which are bounded from above and bounded away from zero, $C_{\mathbb{T}}^{b,0}$, equipped with its Borel σ -algebra, $\mathcal{B}(C_{\mathbb{T}}^{b,0})$. Then

$$\mathbb{G}_n = (C_{\mathbb{T}}^{b,0}, \mathcal{B}(C_{\mathbb{T}}^{b,0}), \{\mathbf{Q}_{\vartheta,n}; \vartheta \in \Theta\}) \quad (4.5)$$

corresponds to the statistical experiment defined in (4.3), cf. Theorem 4.1.

Assume that for any $\vartheta \in \Theta$, there exists a $\vartheta_0 \in \mathbb{R}^d$ and some $h \in \mathbb{R}^d$ such that $\vartheta = \vartheta_0 + n^{-1/2} h$. Consider the Gaussian experiment

$$\mathbb{E} = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \{\mathbf{P}_h; h \in \mathbf{H}\}), \quad (4.6)$$

where $\mathbf{P}_h = \mathcal{N}_p(h, \Gamma_{\vartheta_0}^{-1})$, $\mathbf{H} = \mathbb{R}^d$ with the inner product $\langle x, y \rangle_{\Gamma_{\vartheta_0}} = x^\top \Gamma_{\vartheta_0} y$, and the positive definite matrix Γ_{ϑ_0} is given by

$$\Gamma_{\vartheta_0} = \frac{1}{4\pi} \iint_{\mathbb{T}} (\nabla \log f_{\vartheta_0}) (\nabla \log f_{\vartheta_0})^\top du d\lambda.$$

If $\Theta' \subset \Theta$, then $\mathbb{F}_n(\Theta')$ is the restriction of \mathbb{F}_n over the parameter space Θ' . An analogous notation applies for $\mathbb{G}_n(\Theta')$ and $\mathbb{E}(\Theta')$. In this section we show that these restricted versions of \mathbb{F}_n and \mathbb{G}_n are asymptotically equivalent to the corresponding restricted version of \mathbb{E} over the class of compact subsets of Θ . More precisely, let \mathcal{K}_Θ be the class of compact subsets of Θ ,

Proposition 4.4. *If Assumption 3.1 holds, then*

$$\lim_{n \rightarrow \infty} \Delta(\mathbb{F}_n(K), \mathbb{E}(K)) = 0, \quad (4.7)$$

$$\lim_{n \rightarrow \infty} \Delta(\mathbb{G}_n(K), \mathbb{E}(K)) = 0, \quad (4.8)$$

for any $K \in \mathcal{K}_\Theta$.

Consequently, we can use the triangle inequality and deduce that

$$\lim_{n \rightarrow \infty} \Delta(\mathbb{F}_n(K), \mathbb{G}_n(K)) = 0,$$

for any $K \in \mathcal{K}_\Theta$.

In the following subsections we show the validity of eqs. (4.7) and (4.8).

4.3.1 Local asymptotic equivalence of lsGp

In order to prove the validity of eq. (4.7), we gradually strengthen some well-known results in the literature of statistical inference for locally stationary Gaussian processes.

From Dahlhaus [19] we know that locally stationary Gaussian processes are locally asymptotically normal (LAN). Moreover, under Assumption 3.1 we have proved that this property holds uniformly over compact subsets of Θ , cf. Theorem 3.2 in Ch. 3 of this work.

The stronger type of convergence claimed in (4.7) between the experiments $\mathbb{F}_n(K)$ and $\mathbb{E}(K)$, $K \in \mathcal{K}_\Theta$, can be established from Proposition 3.2 and an application of *uniform λ -convergence of statistical experiments*. An introduction to λ -convergence is given in Section 4.5, at this point, however, we can try and explain some sufficient conditions for an experiment to be uniformly λ -convergent.

The limit experiment $\mathbb{E} = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \{\mathbf{P}_h; h \in \mathbf{H}\})$ is required to be dominated, canonical and \mathcal{K}_Θ -regular. More precisely, the class $\{\mathbf{P}_h; h \in \mathbf{H}\}$ is canonical by definition because it is dominated by \mathbf{P}_0 and the measurable space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is complete and separable. Let $h \in K \subset \mathcal{K}_\Theta$, $\mathcal{Z}_h(x) = d\mathbf{P}_h/d\mathbf{P}_0(x)$ is given by

$$\mathcal{Z}_h(x) = \exp \left\{ \langle h, x \rangle_{\Gamma_{\vartheta_0}} - \frac{1}{2} \|h\|_{\Gamma_{\vartheta_0}}^2 \right\}. \quad (4.9)$$

Observe that $\mathcal{Z}_h(x)$ is continuous in $(x, h) \in \mathbb{R}^d \times K$. The latter is a sufficient condition for \mathbb{E} to be \mathcal{K}_Θ -regular; cf. Section 4.5.

Proof of (4.7). In order to prove this proposition we utilize Theorem 4.15 which is presented in Section 4.5. Since we have already established that the limit experiment \mathbb{E} is a \mathcal{K}_Θ -regular experiment, it remains to show that there exist $\mathcal{B}(\mathbb{R}^n) \setminus \mathcal{B}(\mathbb{R}^d)$ -measurable elements $\lambda^n : \mathcal{B}(\mathbb{R}^n) \rightarrow \mathcal{B}(\mathbb{R}^d)$ such that for any $\varepsilon > 0$ and for every $K \in \mathcal{K}_\Theta$ the following holds

$$\mathcal{L}(\lambda^n \mid \mathbf{P}_{\vartheta,n}) \Rightarrow \mathcal{L}(\lambda \mid \mathbf{P}_{\vartheta_0}) \quad (4.10)$$

$$\sup_{h \in K} \mathbf{P}_{\vartheta_0,n} \left\{ |\mathcal{Z}_{h,n} - \mathcal{Z}_h(\lambda^n)| > \varepsilon \right\} \rightarrow 0, \quad (4.11)$$

where for any $h \in K$, $\mathcal{Z}_{h,n}(X) = d\mathbf{P}_{h,n}/d\mathbf{P}_{\vartheta_0,n}(X)$ for some $\vartheta_0 \in K$, and $\mathcal{Z}_h(X)$ is given in (4.9). Since K is compact it suffices to verify these conditions for convergent sequences. Let $\vartheta_n \rightarrow \vartheta_0 \in K$. Set $\lambda^n = -\sqrt{n}\Gamma_{\vartheta_0}^{-1}\nabla\mathcal{L}_n(\vartheta_n)$ and observe that eq. (3.7) of Proposition 3.2 yields that $\lambda^n \Rightarrow \lambda$ where $\mathfrak{L}(\lambda \mid \mathbf{P}_{\vartheta_0}) = \mathcal{N}_p(0, \Gamma_{\vartheta_0}^{-1})$. This shows that (4.10) holds. Then, note that under the reparametrization $\vartheta_n = \vartheta_0 + n^{-1/2}h_n$ for some $h_n \rightarrow h \in K$, eq. (3.6) yields: $\mathcal{Z}_{\vartheta_n,n}(X) = \exp\{\langle h, \lambda^n \rangle_{\Gamma_{\vartheta_0}} - (1/2)\|h\|_{\Gamma_{\vartheta_0}}^2 + \Psi_n(\vartheta_n^*)\}$, where $\Psi_n(\vartheta_n^*) = -(1/2)\|h\|_{A_n(\vartheta_n^*)}^2$, $A_n(\vartheta_n^*) = \nabla^2\mathcal{L}_n(\vartheta_n^*) - \Gamma_{\vartheta_0}$, with $\|\vartheta_n^* - \vartheta_0\|_2 \leq n^{-1/2}$; cf. proof of Proposition 3.2. Observe that we have made a transition from a ϑ_n -model to an h_n -model, i.e., the parameter of interest is now the vector h_n ; this change depends on the value of ϑ_0 . Utilizing eq. (4.9) we obtain that

$$\mathcal{Z}_{\vartheta_n,n} - \mathcal{Z}_h(\lambda^n) = \exp\left\{\langle h, \lambda^n \rangle_{\Gamma_{\vartheta_0}} - \frac{1}{2}\|h\|_{\Gamma_{\vartheta_0}}^2\right\} (\exp\{\Psi_n(\vartheta_n^*)\} - 1).$$

Again, an application of eq. (3.7) yields that the first term in the right-hand side of this equation converges weakly to some random variable. Since $A_n(\vartheta_n^*) = \nabla^2\mathcal{L}_n(\vartheta_n^*) - \nabla^2\mathcal{L}_n(\vartheta_0) + \nabla^2\mathcal{L}_n(\vartheta_0) - \Gamma_{\vartheta_0} \rightarrow 0$ in $\mathbf{P}_{\vartheta_0,n}$ -probability, cf. Lemma 3.16 and Lemma 3.17, the term in parentheses tends to zero in $\mathbf{P}_{\vartheta_0,n}$ -probability. Consequently, eq. (4.11) holds and completes the proof. \square

4.3.2 Local asymptotic equivalence of a bidimensional white noise model

In this section we demonstrate the validity of eq. (4.8) of Theorem 4.1, i.e., we show that the experiment \mathbb{G}_n converges, with respect to the Δ -pseudodistance, to \mathbb{E} . Recall that the experiment \mathbb{G}_n is given by an observation of the continuous model

$$dz_{(u,\lambda)} = \log f_{\vartheta}(u, \lambda) du d\lambda + \frac{2\sqrt{\pi}}{\sqrt{n}} d\mathbf{W}_{(u,\lambda)}, \quad (u, \lambda) \in \mathbb{T}, \quad (4.12)$$

where $f_\vartheta \in \mathcal{F}_\Theta$ and W is a Brownian motion on $\mathbb{T} = [-\pi, \pi] \times [0, 1]$.

Our claim is shown in several steps. We begin by presenting a series of experiments which are locally asymptotically equivalent to \mathbb{G}_n . Then, we utilize the simplest among these experiments to obtain the LAN property of the bidimensional white noise model given in (4.12). In analogy with our work in the previous section, we strengthen this property: we show first the uniform local asymptotic normality (ULAN) property of this experiment, and then we utilize Theorem 4.15, included in Section 4.5, to establish that \mathbb{G}_n locally asymptotically converges to the experiment \mathbb{E} over \mathcal{K}_Θ .

Equivalently eq. (4.12) can be written as

$$z_{(u,\lambda)} = \int_0^u \int_{-\pi}^\lambda \log f_\vartheta(s, t) ds dt + \frac{2\sqrt{\pi}}{\sqrt{n}} \mathbf{W}_{(u,\lambda)}, \quad (u, \lambda) \in \mathbb{T}.$$

Applying Girsanov's theorem, cf. Appendix II of Ibragimov and Has'minskii [32], we get that $z_{(u,\lambda)}$ has the following probability law:

$$z_{(u,\lambda)} \sim \mathcal{N}\left(\int_0^u \int_{-\pi}^\lambda \log f_\vartheta(s, t) ds dt, \frac{4\pi}{n} u(\lambda + \pi)\right).$$

We localize the parameter $\vartheta \in \Theta$, i.e., we write $\vartheta = \vartheta_0 + h/\sqrt{n}$ for some $h \in \mathbf{H}$ and a suitable vector $\vartheta_0 \in \mathbb{R}^d$. Recall that $\mathbf{H} := \mathbb{R}^d$ with inner product $\langle x, y \rangle_{\Gamma_{\vartheta_0}} = x^\top \Gamma_{\vartheta_0} y$. From Assumption 3.1.2 we deduce that for any h , $\|h\|_{\nabla^2 \log f_{\vartheta_0}}^2$ is finite. Thus, the following Taylor approximation for $\log f_{(\cdot)}(u, \lambda)$ around the point ϑ_0 is valid:

$$\log f_\vartheta(u, \lambda) = \log f_{\vartheta_0}(u, \lambda) + n^{-1/2} (h^\top \nabla \log f_{\vartheta_0}(u, \lambda)) + o(n^{-1}),$$

where ∇ and ∇^2 denote gradient and Hessian, respectively.

We now define the experiment $\mathbb{G}_{n,1}(h)$ given by observations

$$\begin{aligned} dz_{(u,\lambda)} &:= dz_{(u,\lambda)}^{(1)} = (\log f_{\vartheta_0}(u, \lambda)) du d\lambda \\ &+ n^{-1/2} (h^\top \nabla \log f_{\vartheta_0}(u, \lambda)) du d\lambda + \frac{2\sqrt{\pi}}{\sqrt{n}} d\mathbf{W}_{(u,\lambda)}, \quad (u, \lambda) \in \mathbb{T}. \end{aligned} \quad (4.13)$$

The probability law of $Z^{(1)}$ is Gaussian with mean equal to

$$\int_0^u \int_{-\pi}^\lambda \log f_{\vartheta_0}(s, t) ds dt + \frac{1}{\sqrt{n}} \left(\int_0^u \int_{-\pi}^\lambda h^\top \nabla \log f_{\vartheta_0}(s, t) ds dt \right)$$

and variance $(4\pi/n) u (\lambda + \pi)$.

It is known that the square Hellinger distance between Gaussian distributions is given by

$$H^2(\mathcal{N}(\mu_1, \sigma_1), \mathcal{N}(\mu_2, \sigma_2)) = 2 \left[1 - \left(\frac{2\sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2} \right)^{1/2} \exp \left\{ -\frac{(\mu_1 - \mu_2)^2}{4(\sigma_1^2 + \sigma_2^2)} \right\} \right]. \quad (4.14)$$

See eq. (3.7) in Brown and Low [9]. Consequently,

$$\sup_{u, \lambda} H^2(Z(u, \lambda), Z^{(1)}(u, \lambda)) \leq 2 \left[1 - \exp \left\{ -\frac{1}{16} o(n^{-1}) \right\} \right] = o(n^{-1}).$$

Hence, the experiments \mathbb{G}_n and $\mathbb{G}_{n,1}(h)$ are locally asymptotically equivalent over sets of the form $\{\nu \mid \|\nu - \vartheta\| \leq Cn^{-1/2}\}$.

Observe that the equivalence class of $\mathbb{G}_{n,1}(h)$ does not change when the term $\log f_{\vartheta_0}$ is omitted, since this term does not depend on the parameter ϑ , and omitting it signifies a translation of the observed process $Z^{(1)}$ by a known quantity. Therefore, the experiments \mathbb{G}_n and $\mathbb{G}_{n,1}(h)$ are locally asymptotically equivalent to the experiment $\mathbb{G}_{n,2}(h)$ given by observations

$$dz_{(u, \lambda)}^{(2)} = n^{-1/2} (h^\top \nabla \log f_{\vartheta_0}(u, \lambda)) du d\lambda + \frac{2\sqrt{\pi}}{\sqrt{n}} d\mathbf{W}_{(u, \lambda)}. \quad (4.15)$$

We summarize these results in the following

Theorem 4.5. *Suppose that Assumption 3.1 holds. Then, the experiments given by observations*

$$\begin{aligned} dz_{(u, \lambda)} &= (\log f_{\vartheta}(u, \lambda)) du d\lambda + \frac{2\sqrt{\pi}}{\sqrt{n}} d\mathbf{W}_{(u, \lambda)}, \quad (u, \lambda) \in \mathbb{T} \\ dz_{(u, \lambda)}^{(1)} &= (\log f_{\vartheta_0}(u, \lambda)) du d\lambda + \frac{1}{\sqrt{n}} (h^\top \nabla \log f_{\vartheta_0}(u, \lambda)) du d\lambda + \frac{2\sqrt{\pi}}{\sqrt{n}} d\mathbf{W}_{(u, \lambda)}, \quad (u, \lambda) \in \mathbb{T} \\ dz_{(u, \lambda)}^{(2)} &= \frac{1}{\sqrt{n}} (h^\top \nabla \log f_{\vartheta_0}(u, \lambda)) du d\lambda + \frac{2\sqrt{\pi}}{\sqrt{n}} d\mathbf{W}_{(u, \lambda)}, \quad (u, \lambda) \in \mathbb{T} \end{aligned}$$

are locally asymptotically equivalent in the sense of LeCam over sets of the form $\{\vartheta \in \Theta \mid \|\vartheta - \vartheta_0\| \leq Cn^{-1/2}\}$.

Observe that multiplying the observed process Z by a known constant, say \sqrt{n} , signifies re-scaling it and the equivalence class of $\mathbb{G}_{n,2}(h)$ will not be changed. This principle allows us to obtain another equivalent experiment. Namely, the experiment $\mathbb{G}_3(h)$ given by observations

$$dZ_{(u,\lambda)}^{(3)} = h^\top \nabla \log f_{\vartheta_0}(u, \lambda) du d\lambda + 2\sqrt{\pi} d\mathbf{W}_{(u,\lambda)}, \quad (4.16)$$

is equivalent to the experiment $\mathbb{G}_{n,2}(h)$.

For given $n \in \mathbb{N}$, define

$$s_{i,j}^n = n^2 \int_{\frac{i-1}{n}}^{\frac{i}{n}} \int_{-\pi + \frac{2\pi(j-1)}{n}}^{-\pi + \frac{2\pi j}{n}} (\nabla \log f_{\vartheta_0}(u, \lambda)) dZ_{(u,\lambda)}^{(3)}.$$

The variables $\{s_{i,j}^n : i, j = 1, \dots, n\}$ are sufficient for $\{Z_{(u,\lambda)}^{(3)} : (u, \lambda) \in \mathbb{T}\}$. Sufficiency is another principle often used to show equivalence between experiments. Let \mathbb{E}_1 be an experiment with complete and separable parameter space given by observations Y and let \mathbb{E}_2 be an experiment given by observations $T(Y)$. If T is a sufficient statistic w.r.t. \mathbb{E}_1 , then $\Delta(\mathbb{E}_2, \mathbb{E}_1) = 0$; cf. Lemma 3.2 of Brown and Low [9]. Thus the local experiment $\mathbb{G}_4(h)$ given by observations

$$\begin{aligned} Y &:= \iint_{\mathbb{T}} (\nabla \log f_{\vartheta_0}(u, \lambda)) dZ_{(u,\lambda)}^{(3)} \\ &= \left(\iint_{\mathbb{T}} (\nabla \log f_{\vartheta_0}(u, \lambda)) (\nabla \log f_{\vartheta_0}(u, \lambda))^\top du d\lambda \right) h \\ &\quad + 2\sqrt{\pi} \int_{\mathbb{T}} \nabla \log f_{\vartheta_0}(u, \lambda) d\mathbf{W}_{(u,\lambda)}, \end{aligned} \quad (4.17)$$

and the experiment $\mathbb{G}_3(h)$ are equivalent, i.e., $\Delta(\mathbb{G}_3(h), \mathbb{G}_4(h); h) = 0$. Observe that

$$Y = n^{-2} \sum_{i,j} s_{i,j}^n.$$

This implies that an equivalent form for the observations in the experiment $\mathbb{G}_4(h)$ is given by

$$Y = 4\pi \Gamma_{\vartheta_0} h + 4\pi \Gamma_{\vartheta_0}^{1/2} Z, \quad Z \sim \mathcal{N}_d(0, I_{d \times d}),$$

or equivalently,

$$Y := Y^* = h + \Gamma_{\vartheta_0}^{-1/2} Z, \quad Z \sim \mathcal{N}_d(0, I_{d \times d}).$$

This result shows that $Z^{(2)}$, cf. (4.15), is asymptotically normally distributed with mean h and covariance matrix $\Gamma_{\vartheta_0}^{-1}$. Combining the findings in the last paragraph and Theorem 4.5 we have the following:

Corollary 4.6. *Suppose that Assumption 3.1 holds. Then, the bidimensional Gaussian white noise model given by eq. (4.12) is locally asymptotically normal (LAN) with the limit experiment $\mathbb{E} = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \{\mathbf{P}_h = \mathcal{N}(h, \Gamma_{\vartheta_0}^{-1}); h \in \mathbf{H}\})$.*

We now show that this result holds uniformly over \mathcal{K}_{Θ} , the class of closed subsets of Θ . Namely, we establish that the bidimensional white noise model is ULAN over \mathcal{K}_{Θ} . Let $\mathbf{Q}_{\vartheta,n}$ denote the probability law induced by the white noise model with drift $\log f_{\vartheta}$ given in eq. (4.12) of Theorem (4.3).

Proposition 4.7. *Let $K \subset \mathcal{K}_{\Theta}$ and suppose that Assumption 3.1 holds. Then, for any sequence of points $\vartheta_n \in K$ and any sequence of points $u_n \rightarrow u \in \mathbb{R}^d$ there exists a point $\vartheta_0 \in K$ such that the following representation is valid:*

$$\log \frac{d\mathbf{Q}_{\vartheta_n + \varphi_n(\vartheta_n)u_n, n}}{d\mathbf{Q}_{\vartheta_n, n}}(Z) - \langle u, \xi \rangle_{\Gamma_{\vartheta_0}} + \frac{1}{2} \|u\|_{\Gamma_{\vartheta_0}}^2 \xrightarrow{\mathbf{Q}_{\vartheta_n, n}} 0, \quad (4.18)$$

where $\xi \sim \mathcal{N}(0, I_{d \times d})$, $\varphi_n(\vartheta_n) = n^{-1/2} I_{d \times d}$.

Proof. This result is a consequence of the previous calculations. Namely, the observations

$$dz_{1,n} = \log f_{\vartheta_n + \varphi(\vartheta_n)u_n}(u, \lambda) du d\lambda + \frac{2\sqrt{\pi}}{\sqrt{n}} d\mathbf{W}(u, \lambda)$$

are asymptotically equivalent to observations

$$dz_{2,n} = \frac{u_n^\top}{\sqrt{n}} \nabla \log f_{\vartheta_n}(u, \lambda) du d\lambda + \frac{2\sqrt{\pi}}{\sqrt{n}} d\mathbf{W}(u, \lambda),$$

which in turn may be written as $Y^* = u + \Gamma_{\vartheta_0}^{-1/2} \xi$ provided $\vartheta_n \rightarrow \vartheta_0 \in K$. Here $\xi \sim \mathcal{N}(0, I_p)$. Therefore, the law induced by observations $z_{1,n}$ is asymptotically a normal with mean u and covariance matrix $\Gamma_{\vartheta_0}^{-1}$.

Analogously, the observations

$$dZ_{1,n}^\circ = \log f_{\vartheta_n}(u, \lambda) du d\lambda + \frac{2\sqrt{\pi}}{\sqrt{n}} d\mathbf{W}(u, \lambda)$$

asymptotically induce a $\mathcal{N}(0, \Gamma_{\vartheta_0}^{-1})$ law.

In view of these results and since,

$$\frac{\mathbf{Q}_{\vartheta_n + \varphi(\vartheta_n)u_n, n}}{\mathbf{Q}_{\vartheta_n, n}}(Z) = \exp \left\{ -\frac{1}{2} \left(-2\langle u_n, \xi \rangle_{\Gamma_{\vartheta_n}} + \|u_n\|_{\Gamma_{\vartheta_n}}^2 \right) \right\},$$

we deduce that

$$\log \frac{\mathbf{Q}_{\vartheta_n + \varphi(\vartheta_n)u_n, n}}{\mathbf{Q}_{\vartheta_n, n}}(Z) - \langle u, \xi \rangle_{\Gamma_{\vartheta_0}} + \frac{1}{2} \|u\|_{\Gamma_{\vartheta_0}}^2 \xrightarrow{\mathbf{Q}_{\vartheta_n, n}} 0.$$

This completes the proof. \square

Proof of (4.8). In view of Proposition 4.18, we can mimick the steps used in the proof of eq. (4.7) to obtain this result. The details are omitted. \square

4.4 Preliminary estimators

Under localization, i.e., $\vartheta = \vartheta_0 + h/\sqrt{n}$, for some $h \in \mathbb{R}^p$, Dahlhaus [19] showed the asymptotic normality and efficiency of the maximum likelihood estimator, $\widehat{\vartheta}_n$, for the true parameter, ϑ_0 , of a locally stationary Gaussian process; cf. Theorem 2.1 and Theorem 4.2 of [19], respectively.

From the results of Chapter 3 we know that there is a family of Bayesian estimators which are consistent, asymptotic normal and efficiency for ϑ , and this result holds uniformly for ϑ , cf. Theorem 3.5. Similar results are valid for the maximum likelihood estimator of ϑ within the experiment $\{\mathbf{P}_{\vartheta,n} = \mathcal{N}(0, \Sigma_\vartheta)\}$ from a lsGp, cf. Theorem 3.8.

As an application of these results, Corollary 3.7 yields that for any $K \in \mathcal{K}_\Theta$ there exists a number τ large enough such that for any $\varepsilon > 0$ the following inequality

$$\sup_{\vartheta \in K} \mathbf{P}_\vartheta \left\{ \sqrt{n} \|\tilde{\vartheta}_n - \vartheta\|_2 > \tau \right\} \leq \varepsilon.$$

is valid. A similar result holds for the maximum likelihood estimator.

In this section, we show that a similar result holds within the white noise problem

$$dz_{(u,\lambda)} = \log f_\vartheta(u, \lambda) du d\lambda + \frac{2\sqrt{\pi}}{\sqrt{n}} d\mathbf{W}_{(u,\lambda)}, \quad (u, \lambda) \in \mathbb{T}.$$

Again, we utilize the estimation method developed by Ibragimov and Hansminkii [32]. Below we verify that the conditions of Theorem 3.1.2 of [32] are satisfied. We present these conditions as a series of lemmas.

4.4.1 Estimators in a bidimensional white noise model

In this section we establish the existence of uniform local asymptotic minimax estimators for the drift function of the white noise model given in the experiment eq. (4.12). In order to obtain our result, once again we utilize Theorem 3.1.2 of Ibragimov and Hasmiskii [32].

Throughout this section, we use the convention that $\mathbf{E}_{\vartheta,n} := \mathbf{E}_{\mathbf{Q}_{\vartheta,n}}$ where $\mathbf{Q}_{\vartheta,n}$ is defined in (4.5); \mathcal{K}_Θ is the class of compact sets of Θ ; $\|x\|_2$ is the Euclidean norm of $x \in \mathbb{R}^d$ and for any $d \times d$ matrix M , $\|M\| = (\max. \text{characteristic root of } M^* M)^{1/2}$ where M^* denotes the conjugate transpose of M .

Lemma 4.8 (Condition N1 in [32]). *For some nondegenerate matrix $\varphi_n(\vartheta)$, for any set $K \in \mathcal{K}_\Theta$ and arbitrary sequences $\vartheta_n \in K$, $\vartheta_n + \varphi(\vartheta_n)u_n \in K$, $u_n \rightarrow u$, $n \rightarrow \infty$, the representation*

$$\begin{aligned}\tilde{\mathcal{Z}}_{n,\vartheta_n}(u_n) &= \frac{d\mathbf{Q}_{\vartheta_n+\varphi(\vartheta_n)u_n,n}}{d\mathbf{Q}_{\vartheta_n,n}}(Z) \\ &= \exp \left\{ \langle \xi, u \rangle - \frac{1}{2} \|u\|^2 + \psi_n(u_n, \vartheta_n) \right\},\end{aligned}$$

is valid; here $\xi \sim \mathcal{N}(0, I_p)$ and $\psi_n(u_n, \vartheta_n) \rightarrow 0$ in $\mathbf{Q}_{\vartheta_n,n}$ -probability.

Proof. The lemma follows by an application of Proposition 4.7 with $\varphi_n(\vartheta) = n^{-1/2}I_{d \times d}$ and $\psi_n(u_n, \vartheta_n) = 0$. \square

Lemma 4.9 (Condition N2 in [32]). *For any set $K \in \mathcal{K}_\Theta$,*

$$\lim_{n \rightarrow \infty} \sup_{\vartheta \in K} \text{tr} \{ \varphi_n(\vartheta) \varphi_n^\top(\vartheta) \} = 0$$

Proof. It is immediate since $\varphi_n(\vartheta) = n^{-1/2}I_{d \times d}$. \square

Lemma 4.10 (Condition N3 in [32]). *Let $U_n(\vartheta) = \varphi_n^{-1}(\vartheta)(\Theta - \vartheta)$. For any set $K \in \mathcal{K}_\Theta$ and some $\beta > 0$ and $m > 0$, $B = B(K)$, $a = a(K)$,*

$$\sup_{\vartheta \in K} \sup_{\substack{v, w \in U_n(\vartheta) \\ |v| < R, |w| < R}} \|w - v\|_2^{-\beta} \mathbb{E}_\vartheta \left| \tilde{\mathcal{Z}}_{n,\vartheta}^{1/m}(w) - \tilde{\mathcal{Z}}_{n,\vartheta}^{1/m}(v) \right|^m \leq B(1 + R^a).$$

Proof. Again, we utilize Lemma 1.1 in [32]. Let $I_n(\vartheta)$ denote the Fisher's information matrix of the white noise model given in eq. (4.5). Since the regularity condition on the model (4.5) can be established in analogy with Lemma 3.13, we omit the details. It suffices to show the boundedness of

$$\sup_{\vartheta \in K} \sup_{\|u\| < R, \vartheta + u \in \Theta} \left\| I_n^{-1/2}(\vartheta) I_n(\vartheta + u) I_n^{-1/2}(\vartheta) \right\|.$$

Since K is compact, we focus on convergent sequences $\vartheta_n \rightarrow \vartheta_0 \in K$. As argued above the model $dZ_{2,n} = \log f_{\vartheta_n}(u, \lambda) du d\lambda + 2\sqrt{\pi}n^{-1/2} d\mathbf{W}(u, \lambda)$ asymptotically induces

a $\mathcal{N}(0, \Gamma_{\vartheta_0}^{-1})$ law. We deduce that the Fisher's information of this model $I_n(\vartheta_n) \rightarrow \Gamma_{\vartheta_0}$ provided $\vartheta_n \rightarrow \vartheta_0$.

Consequently, after using Proposition V 1.8 and Theorem X 1.1 of Bhatia [5],

$$\begin{aligned} \|I_n^{-1/2}(\vartheta_n) I_n(\vartheta_n + u_n) I_n^{-1/2}(\vartheta_n)\| &\leq \|I_n^{-1}(\vartheta_n)\| \|I_n(\vartheta_n + u_n)\| \\ &\leq \left[\|\Gamma_{\vartheta_0}^{-1}\| \|I_n(\vartheta_n) - \Gamma_{\vartheta_0}\| \|I_n^{-1}(\vartheta_n)\| + \|\Gamma_{\vartheta_0}^{-1}\| \right] \\ &\quad \left[\|I_n(\vartheta_n + u_n) - \Gamma_{\vartheta_0+u}\| + \|\Gamma_{\vartheta_0+u}\| \right]. \end{aligned}$$

From Lemma A4 of Dahlhaus [19] follows that $\|\Gamma_{\vartheta}^{-1}\|$ is uniformly bounded by constants. The boundedness of $\|I_n^{-1}(\vartheta)\|$ can be established similarly as in the proof of Lemma 3.13. Since $I_n(\vartheta_n) \rightarrow \Gamma_{\vartheta_0}$ provided $\vartheta_n \rightarrow \vartheta_0$, the above quantity is bounded for n large enough. This completes the proof. \square

Lemma 4.11 (Condition N4 in [32]). *For any set $K \in \mathcal{K}_{\Theta}$ and any $N > 0$, there exists an $n_0(N, K)$ such that*

$$\sup_{\vartheta \in K} \sup_{n_0 < N} \sup_{u \in U_n(\vartheta)} \|u\|_2^N \mathbb{E}_{\vartheta} \widetilde{\mathcal{Z}}_{n,\vartheta}^{1/2}(w) < \infty.$$

Proof. The compactness of K allows us to focus on convergent sequences $\vartheta_n \rightarrow \vartheta_0 \in K$ and $u_n \rightarrow u \in \mathbb{R}^p$. From Lemma 4.8 we know that for n large enough, $\widetilde{\mathcal{Z}}_n^{1/2}(u_n)$ may be written as

$$\widetilde{\mathcal{Z}}_n^{1/2}(u_n) = \exp \left\{ \frac{1}{2} \langle \xi, u \rangle_{\Gamma_{\vartheta_0}} - \frac{1}{4} \|u\|_{\Gamma_{\vartheta_0}}^2 + \varphi(u_n, \vartheta_n) \right\},$$

where $\xi \sim \mathcal{N}(0, I_p)$ and $\varphi(u_n, \vartheta_n) \rightarrow 0$ in $\mathbf{Q}_{\vartheta_n, n}$ -probability. Thus $\mathbb{E}_{\vartheta_n, n} \widetilde{\mathcal{Z}}_n^{1/2}(\vartheta_n)$ is finite. We complete the proof by noting that if $u_n \in U_n(\vartheta_n)$, $\|u_n\|^N \leq \text{diam}(\Theta)^N$ which is finite since Θ is compact. \square

These results allows us to use Theorem 3.1.2 of Ibragimov and Hasminskii and conclude that there are suitable estimates for the parameter ϑ in the model (4.5). In analogy

with the discussion following Proposition 3.8 and Corollary 3.7 we state the main result of this section.

Corollary 4.12. *Suppose that Assumption 3.1 holds. Then, in the model (4.5) there exists a number τ large enough such that for any $\varepsilon > 0$ the following inequality*

$$\sup_{\vartheta \in \Theta} \mathbf{Q}_{\vartheta} \left\{ \sqrt{n} \|\widehat{\vartheta}_n - \vartheta\| > \tau \right\} \leq \varepsilon,$$

is valid.

4.5 Strong local convergence of statistical experiments

We have used uniform λ -convergence to obtain the local asymptotic equivalence of the experiments introduced in the main theorem of this chapter. In this section, we present a brief introduction of this type of convergence of statistical experiments. For a more detailed treatment of λ -convergence the reader is referred to Shiryaev and Spokoiny [43]. We begin with some definitions.

An experiment $\mathbb{E} = (\Omega, \mathcal{A}, \{\mathbf{P}_{\vartheta}; \vartheta \in \Theta\})$ is said to be dominated if there is a σ -finite measure μ on (Ω, \mathcal{A}) such that all the measures \mathbf{P}_{ϑ} , $\vartheta \in \Theta$, are absolutely continuous with respect to μ . The experiment \mathbb{E} will be called canonical if the space (Ω, \mathcal{A}) is complete and separable. If the family $\{\mathbf{P}_{\vartheta} : \vartheta \in \Theta\}$ is dominated by μ , it is possible to choose the dominating measure μ in the following form

$$\mu = \sum_{i=1}^{\infty} c_i \mathbf{P}_{\vartheta_i}, \quad c_i \geq 0, \quad \sum_{i=1}^{\infty} c_i = 1, \quad \vartheta_i \in \Theta. \quad (4.19)$$

See Strasser [44], p. 94.

Suppose that the statistical experiments $\mathbb{E}_n = (\Omega^n, \mathcal{A}^n, \{\mathbf{P}_{\vartheta}^n; \vartheta \in \Theta\})$ are given such

that $\Theta^n \uparrow \Theta$. Set $S_i = \{i : \vartheta_i \in \Theta^n \text{ satisfying (4.19)}\}$ and denote

$$\mu^n = \frac{\sum_{S_i} c_i \mathbf{P}_{\vartheta_i}^n}{\sum_{S_i} c_i}. \quad (4.20)$$

Let \mathcal{K} be a collection of subsets of Θ .

Definition 4.13. Let $\mathbb{E} = (\Omega, \mathcal{A}; \{\mathbf{P}_{\vartheta}, \vartheta \in \Theta\}, \mu)$ be a dominated canonical experiment. Let the experiments $\mathbb{E}_n = (\Omega^n, \mathcal{A}^n; \{\mathbf{P}_{\vartheta}^n, \vartheta \in \Theta^n\})$ be given such that $\Theta^n \uparrow \Theta$. The experiments \mathbb{E}_n are said to \mathcal{K} -uniformly λ -converge to \mathbb{E} , in notation $\mathbb{E}_n \xrightarrow{\lambda(\mathcal{K})} \mathbb{E}$, if there exist $\mathcal{A}^n/\mathcal{A}$ -measurable elements $\lambda^n = \lambda^n(\omega) : \Omega^n \rightarrow \Omega$ such that for any $\varepsilon > 0$ and for every $K \in \mathcal{K}$ the following conditions are satisfied:

$$\mathfrak{L}(\lambda^n | \mu^n) \Rightarrow \mathfrak{L}(\lambda | \mu), \quad (4.21)$$

$$\sup_{\vartheta \in K} \mu^n \left\{ |\mathfrak{Z}_{\vartheta}^n - \mathfrak{Z}_{\vartheta}(\lambda^n)| > \varepsilon \right\} \rightarrow 0, \quad (4.22)$$

where μ^n is given in (4.20), $\mathfrak{Z}_{\vartheta}(x) = \frac{d\mathbf{P}_{\vartheta}}{d\mu}(x)$ is the Radon-Nikodym derivative of \mathbf{P}_{ϑ} w.r.t. μ , and $\mathfrak{Z}_{\vartheta}^n(x) = \frac{d\mathbf{P}_{\vartheta}^n}{d\mu^n}(x)$.

Remark 4.14. If μ^n is mutually absolutely continuous w.r.t. the family $\{\mathbf{P}_{\vartheta}^n; \vartheta \in \Theta\}$, the equation (4.22) holds if for any $\varepsilon > 0$ and for any $K \in \mathcal{K}$,

$$\sup_{\vartheta \in K} \mathbf{P}_{\vartheta}^n \left\{ |\mathfrak{Z}_{\vartheta}^n - \mathfrak{Z}_{\vartheta}(\lambda^n)| > \varepsilon \right\} \rightarrow 0.$$

If \mathbb{E} is a Gaussian shift experiment with $\mathbf{P}_{\vartheta} \sim \mathcal{N}(\vartheta, I_{d \times d})$, $\vartheta \in \mathbb{R}^d$ and $\mathfrak{Z}_{\vartheta}(x) = \exp\{\langle \vartheta, x \rangle - \|\vartheta\|^2/2\}$, then (4.21) and (4.22) are the conditions for asymptotic normality around some point $\vartheta_0 \in \mathbb{R}^d$ with $\mathbf{P} = \mathbf{P}_{\vartheta_0}$ and $\mathbf{P}^n = \mathbf{P}_{\vartheta_0}^n$. In view of this, uniform λ -convergence generalizes LeCam's uniform asymptotic normality (UAN) theory.

Under some regularity conditions on the likelihood functions \mathfrak{Z}_{ϑ} an UAN property can be strengthened and yields a local asymptotic equivalence. For instance, it is necessary that for any $K \in \mathcal{K}$ the family $\{\mathfrak{Z}_{\vartheta}; \vartheta \in K\}$ be uniformly integrable under the

measure μ as well as uniformly continuous with respect to the measure μ . If $\{\mathcal{Z}_\vartheta; \vartheta \in K\}$ meets these two properties then \mathbb{E} is said to be \mathcal{K} -regular. When Θ is a metric space and \mathcal{K} is the class of its compact sets then, a sufficient condition for \mathcal{K} -regularity is that the density functions $\mathcal{Z}_\vartheta(x) = \frac{d\mathbf{P}_\vartheta}{d\mu}(x)$ be continuous in (x, ϑ) on $\Omega \times \Theta$, cf. Lemma 2.3 of Shiryaev and Spokoiny [43].

We now present the main theorem of this section. Let Δ be the Le Cam's pseudodistance between experiments having the same parameter space.

Theorem 4.15. *Let \mathbb{E} be a \mathcal{K} -regular experiment such that $\mathbb{E}_n \xrightarrow{\lambda(\mathcal{K})} \mathbb{E}$. Then for any $K \in \mathcal{K}$ the uniform strong convergence, in notation $\mathbb{E}_n \xrightarrow{\Delta(\mathcal{K})} \mathbb{E}$, holds. That is, for each $K \in \mathcal{K}$*

$$\Delta(\mathbb{E}_n, \mathbb{E}; K) \rightarrow 0, \quad n \rightarrow \infty.$$

Proof. See Theorem 2.9 in [43]. □

BIBLIOGRAPHY

- [1] AUDENAERT, K., NUSSBAUM, M., SZKOLA, A., AND VERSTRAETE, F. Asymptotic error rates in Quantum hypothesis testing. *Communications in Mathematical Physics* 278 (2008), 251–283.
- [2] BAHADUR, R. R. *Some limit theorems in Statistics*, vol. 4 of *Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, 1971.
- [3] BARRON, A. Entropy and the central limit theorem. *The Annals of Probability* 14 (1986), 336–342.
- [4] BERAN, J. On parameter estimation for locally stationary long-memory process. *Journal of Statistical Planning and Inference* 139 (2009), 900–915.
- [5] BHATIA, R. *Matrix Analysis*. Springer-Verlag, New York, 1997.
- [6] BILLINGSLEY, P. *Convergence of Probability Measures*. Wiley, New York, 1968.
- [7] BLAHUT, R. E. Hypothesis testing and information theory. *IEEE Trans. Inf. Theory* 20 (1974), 405–417.
- [8] BOUAZIZ, M. Testing Gaussian sequences and asymptotic inversion of Toeplitz operators. *Probability and Mathematical Statistics* 14(2) (1993), 207–222.
- [9] BROWN, L. D., AND LOW, M. Asymptotic equivalence of nonparametric regression and white noise. *The Annals of Statistics* 24(6) (1996), 2384–2398.
- [10] BUCKLEW, J. *Large deviations techniques in decision, simulation, and estimation*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1990.
- [11] CHERNOFF, H. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics* 23(4) (1952), 493–507.
- [12] CHERNOFF, H. Large-sample theory: Parametric case. *The Annals of Mathematical Statistics* 27(1) (1956), 1–22.
- [13] COURSOL, J., AND DACUNHA-CASTELLE, D. Sur la formule de Chernoff pour deux processus gaussiens stationnaires. *C. R. Acad. Sci. Paris* 288 A (1979), 769–770.

- [14] CSISZÁR, I., AND LONGO, G. On the error exponent for source coding and for testing simple statistical hypotheses. *Studia Sci. Math. Hungarica* 6 (1971), 181–191.
- [15] DACHIAN, S. On the limiting likelihood ratio processes of some change-point type models. *Journal of Statistical Planning and Inference* 140 (2010), 2682–2692.
- [16] DACHIAN, S. Estimation of the location of a 0-type or ∞ -type singularity by poisson observations. *Statistics* 45(5) (2011), 509–523.
- [17] DACHIAN, S., AND NEGRI, I. On compound poisson processes arising in change-point type models. *Statistical Inference for Stochastic Processes* 14 (2011), 255–271.
- [18] DAHLHAUS, R. On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Process. Appl.* 62 (1996a), 139–168.
- [19] DAHLHAUS, R. Maximum likelihood estimation and model selection for locally stationary processes. *Nonparametric Statistics* 6 (1996b), 171–191.
- [20] DAHLHAUS, R. Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In *Athens Conference on Applied Probability and Time Series*, vol. 2. Springer, New York, 1996c, pp. 145–159. P. M. Robinson and M. Rosenblatt (Editors).
- [21] DAHLHAUS, R. Local inference for locally stationary time series based on the empirical spectral measure. *Journal of Econometrics* 151 (2009), 101–112.
- [22] DAHLHAUS, R., AND POLONIK, W. Empirical spectral processes for locally stationary time series. *Bernoulli* 15 (2009), 1–39.
- [23] DAVIES, R. Asymptotic inference in stationary Gaussian time series. *Adv. Appl. Probab.* 5 (1973), 469–497.
- [24] DEMBO, A., AND ZEITOUNI, O. *Large deviations techniques and applications*. Stochastic Modelling and Applied Probability. Springer-Verlag, Heidelberg, 1998.
- [25] ECKLEY, I., NASON, G., AND TRELOAR, R. Locally stationary wavelet fields with application to the modelling and analysis of image texture. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 59 (2010), 595–616.
- [26] FERREIRA, G., RODRÍGUEZ, A., AND LAGOS, B. Kalman filter estimation for a regression model with locally stationary errors. *Computational Statistics and Data Analysis* 62 (2013), 52–69.

- [27] GAPEEV, P., AND KÜCHLER, U. On large deviations in testing Ornstein-Uhlenbeck-type models. *Stat. Infer. Stoch. Process.* 11 (2008), 143–155.
- [28] GENON-CATALOT, V., AND PICARD, D. *Éléments de statistique asymptotique*. Mathématiques & Applications. Springer-Verlag, Paris, France, 1993.
- [29] GOLUBEV, G. K., NUSSBAUM, M., AND ZHOU, H. H. Asymptotic equivalence of spectral density estimation and Gaussian white noise. *Ann. Statist.* 38, 1 (2010), 181–214.
- [30] HIRUKAWA, J., AND TANIGUCHI, M. LAN theorem for non-Gaussian locally stationary processes and its applications. *Journal of Statistical Planning and Inference* 136 (2006), 640–688.
- [31] Hoeffding, W. Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.* 36 (1965), 369–401.
- [32] IBRAGIMOV, I. A., AND HASMINSKII, R. Z. *Statistical estimation: Asymptotic theory*. Applications of Mathematics (16). Springer-Verlag, New York, 1981.
- [33] LE CAM, L. Convergence of estimates under dimensionality restrictions. *The Annals of Statistics* 1(1) (1973), 38–53.
- [34] LE CAM, L. On local and global properties in the theory of asymptotic normality of experiments. *Stochastic Processes and Related Topics* 1 (1975), 13–54.
- [35] LE CAM, L. Sur l’approximation de familles de mesures par des familles Gaussiennes. *Ann. Inst. Henri Poincaré* 21(3) (1985), 225–287.
- [36] LE CAM, L., AND YANG, G. *Asymptotics in Statistics*. Springer Series in Statistics. Springer-Verlag, New York, 2000.
- [37] MAHARAJ, E. A., AND ALONSO, A. M. Discrimination of locally stationary time series using wavelets. *Comput. Stat. Data Anal.* 52, 2 (Oct. 2007), 879–895.
- [38] MEISTER, A., AND REISS, M. Asymptotic equivalence for nonparametric regression with non-regular errors. *Probability Theory and Related Fields* (2011), 1–29. 10.1007/s00440-011-0396-x.
- [39] MILLAR, P. The minimax principle in asymptotic statistical theory. In *Ecole d’Eté de Probabilités de Saint-Flour XI-1981*. Springer-Verlag, Berlin-Heidelberg-New York, 1981, pp. 75–267. A. Dolb and B. Eckmann (Editors).

- [40] NUSSBAUM, M. Asymptotic equivalence of density estimation and Gaussian white noise. *The Annals of Statistics* 24 (1996), 2399–2430.
- [41] PALMA, W. On the sample mean of locally stationary long-memory processes. *Journal of Statistical Planning and Inference* 140(12) (2010), 3764–3774.
- [42] PRIESTLY, M. B. *Spectral analysis and time series, Vol. 2*. Monographs and textbooks. Academic Press, London, 1981.
- [43] SHIRYAEV, A. N., AND SPOKOINY, V. G. *Statistical experiments and decisions. Asymptotic theory*. Advanced series on statistical science & applied probability. World Scientific, Singapore, 2000.
- [44] STRASSER, H. *Mathematical Theory of Statistics*. Studies in Mathematics. de Gruyter, Berlin-New York, 1985.
- [45] TANIGUCHI, M. On large deviation asymptotics of some tests in time series. *Journal of Statistical Planning and Inference* 97 (2001), 191–200.
- [46] TANIGUCHI, M., AND KAKIZAWA, Y. *Asymptotic Theory of Statistical Inference for Time Series*. Springer Series in Statistics. Springer-Verlag, New York, 2000.
- [47] UCHIDA, M., AND YOSHIDA, N. adaptive estimation of an ergodic diffusion process based on sample data. *Stochastic processes and their applications* 122 (2012), 2885–2924.
- [48] WALKER, A. M. Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *J. Austral. Math. Soc.* 4 (1964), 363–384.
- [49] ZANI, M. Large deviations for quadratic forms of locally stationary processes. *Journal of Multivariate Analysis* 81 (2002), 205–228.